FACULTY OF ECONOMICS
AND BUSINESS

KU LEUVEN

Essays on Local Housing and Real Estate Brokerage Markets

Dissertation presented to obtain the
Degree of Doctor in Economics

by

Roel Helgers

Number 543

2016

Doctoral committee

Supervisor:

Prof. dr. Erik Buyst, KU Leuven

Members:

Prof. dr. Frank Verboven, KU Leuven
Prof. dr. Maarten Goos, Utrecht University and KU Leuven
Prof. dr. Jan Rouwendal, VU University Amsterdam
Prof. dr. Jan Mutl, EBS University of Business and Law (Wiesbaden)

Chair:

Prof. dr. Erwin Ooghe, KU Leuven

Acknowledgements

Writing this dissertation would not have been possible without the support I received from many persons. Although naming all of them explicitly is impossible, I would nevertheless like to take this opportunity to express my gratitude towards several persons.

First of all, I would like to thank my supervisor Erik Buyst for the excellent guidance in the process of writing this dissertation and his assistance, help and mentorship in general during my PhD. Second of all, I would like to thank Frank Verboven and Maarten Goos with whom I also closely worked together. I believe that the many fruitful discussions we've had over the past couple of years have made me a better researcher and economist. I would also like to express my gratitude towards Jan Rouwendal and Jan Mutl. Not only for their valuable comments and suggestions during the preliminary defense, which I believe have greatly benefited my work, but also more generally for the many interesting discussions we had during seminars and for providing a source of inspiration. I feel truly honored that you are willing to spend so much time and effort, including multiple visits to Leuven, to support my research.

Special thanks furthermore goes out to Bert Willekens, co-author of the last paper in this dissertation. I sincerely believe and feel that bundling our forces over the last couple of years has not only led to a paper which I believe to be is of high quality, but I also really enjoyed working with you and truly feel that our countless discussions have made me a better economist overall. I would further like to thank all of my past and current colleagues, not only for all the stimulating research environment, but also for providing an environment in which I felt comfortable and could develop myself fully. I would especially like to express my gratitude towards Koen Declercq and Simon Miegielsen, with whom I started this venture at our department and shared many highs and depths, and Sven Damen, Frank Vastmans and Geert Goeyvaerts, with whom I discussed and debated on many housing and non-housing related topics on a daily basis in our office. I would also like to thank Karla, Heidi, Pascale and Andras for their excellent administrative support.

I would also like to take this opportunity to thank Iain Cook, Pol Vanacker, Alain Declercq and Claude Darimont of ERA Belgium. You were not only kind enough to let me use the excellent data you gather, which I analysed over the course of several years, but also shared your experience "from the field" and provided another platform to present my results. Thanks furthermore goes out to the people from the Flemish Policy Research Center Housing (Steunpunt Wonen). The countless meetings and seminars and the people I met there were a crucial part of my intellectual development.

Next, I would like to express my utmost gratitude to my family and friends. I am especially grateful to my parents for always being there, for believing in me, and for giving me all the opportunities. I would also like to thank my brother Karel, who has helped me out a lot with a wide variety of difficulties I encountered and has served as a sounding board, and my father-, mother-, and sisters-in-law for their support. Last, but certainly not least, I would like to thank my lovely wife Dana. Without your support, love and perhaps most importantly patience, I would most certainly not have been able to complete this journey. Although finishing this dissertation marks the completion of one phase in my life, I also believe that it just as well signifies the start of a new phase in our lives. I most certainly look forward to what the future has in store for us.

Contents

List of Tables

We Value the Homes of our Neighbors?

IV Intermediation in Markets with Buyer and Seller Selection:

Theory and an Application to Real Estate Brokerage

List of Figures

General Introduction

For many households the decision to purchase a house is one of most important and difficult decisions in their lifetime. As with any durable good, the decision to buy a house does not only entail a consumption dimension, but at the same time is an investment choice. In fact, equity in owner-occupied homes is the dominant form of wealth for most American and European households. It has been documented that changes in housing wealth can have protrusive effects on household consumption (Case *et al.*, 2005) and that developments in housing markets are not just a reflection of economic activity, but are among the driving forces of business cycles (Iacoviello & Neri, 2010). It is perhaps not surprising that the 2001-2006 US housing bubble and its subsequent collapse in 2007, which is generally considered as one of the main causes of the global financial crisis, have renewed academic interest in housing markets and prices and their relation with the rest of the economy. While the studies by Case *et al.* (2005) and Iacoviello & Neri (2010) show that there interactions between developments in housing markets and developments in other markets, the current dissertation focuses mainly on the interactions between different geographical housing markets and interactions between agents in these markets.

Housing is characterized by a peculiar combination of features, which include durability, multidimensional heterogeneity and quality differentiation, indivisibility in consumption, a mix of consumption and investment motives, and the importance of transaction costs and asymmetric information (Quigley, 1979; Arnott, 2013). It is no wonder then that these characteristics have a profound impact on the workings of housing and real estate brokerage markets and play a central role in this dissertation. The aforementioned properties lead Duncan Maclennan (1982) to conclude that: "*the housing market is really not a neoclassical exchange market, but is rather a set of overlapping submarkets differentiated by tenure, location, size and quality.*" Since each house is unique, it can even be argued that every property that is put up for sale constitutes a single market, where a single seller potentially faces zero, one or multiple buyers. Despite that no two properties are exactly alike, they are however substitutable up to a certain degree. Future buyers are likely to visit multiple housing units before they decide to purchase a property. As this process is costly both in terms of money and time, a stable equilibrium might only be achieved over a certain period of time.[1] Since it is also costly for sellers to search for an eligible buyer, due to for example holding- and promotional costs, real estate agents have traditionally acted as intermediaries between sellers and buyers to speed up the matching process. In recent decades, however, the rise of modern information and communication technologies has had a profound impact upon the real estate brokerage industry.

In this dissertation we focus on some of the characteristics that define the housing commodity and their effects on market outcomes. In the next few paragraphs a short summary of each of the chapters presented in this dissertation is provided. In the first chapter we focus on spatial and temporal diffusion of housing prices in Belgium and are especially interested in the role of the linguistic border. In the second chapter we investigate whether housing prices close to the Flemish-Dutch border are higher as a result of spatial arbitrage. In the third chapter, we examine the effects of the housing stock composition at the neighborhood level on housing prices. In the fourth chapter, finally, we start from the observation that high search costs as a result of heterogeneity and spatial fixity of the housing good cause (buyers and) sellers to hire a real estate agent. In this chapter, we focus on the optimal market structure and market efficiency of the real estate brokerage industry.

[1] In *Flash Boys: a Wall Street Revolt* (2014), Michael Lewis, an American non-fiction author and financial journalist who also worked as a bond salesman in the past, focuses on the rise of high-frequency trading in the US equity market. Lewis states that this market "is rigged" by high frequency traders who front run (front running is the unethical practice of a stockbroker executing orders on a security for its own account while taking advantage of advance knowledge of pending orders from its customers.) orders placed by investors. In the first chapter of the book, Lewis describes how Spread Networks' secretive new 827-mile cable running as straight as possible, through mountains and under rivers, from Chicago to New Jersey reduces the journey time for data from 17 to 13 milliseconds. This $ 300 million project was designed to connect the financial markets of Chicago and New York City where one could think front running might happen with a few millisecond advantage. This rather absurd anecdote suggests that arbitrage between geographically differentiated equity markets in the US, unlike housing markets, is instantaneous and without any frictions.

As was already argued, the unique combination of features that underpin the housing commodity and result in substantial search costs, might cause that housing prices in different regions only slowly adjust to a shock. In the UK, but also in other countries, a large academic literature has emerged that investigates regional convergence of - and spatial and temporal diffusion patterns in - housing prices. In the UK, many authors have found that house prices exhibit a distinct spatial pattern over time, rising first in a cyclical upswing in the south-east and then spreading out over the rest of the country (Meen, 1999). This pattern is often referred to as the *ripple effect hypothesis*. Although the statistical evidence is overwhelming and convincing, there are few studies that provide convincing economic explanations. In the **first chapter**, which is joint work with Erik Buyst, we contribute to this literature by examining to what extent the linguistic border that splits Belgium into two distinct regions plays a role in the convergence and spatial and temporal diffusion of housing prices using aggregated data from 22 *a priori* defined Belgian local markets. We believe that this is an interesting exercise since migration, equity transfer and spatial arbitrage, which are potential explanations for the observed ripple effect, are likely to be influenced by the presence of a linguistic border. Methodologically, we contribute to an expanding literature on spatial (macro) panels (e.g. Breitung & Pesaran, 2008). The results presented in this chapter suggest that, similar to the UK, a ripple effect is observed in Belgium, where prices along the axis Antwerp-Brussels-Namur rise first after an initial shock to the dominant region, Antwerp. Subsequently, prices in the peripherally located districts in the east and west of Belgium follow their more centrally located counterparts. Our findings suggest that the linguistic border plays an ambiguous role. While the districts located along the centrally located north-south axis Antwerp-Brussels-Namur are highly integrated despite the linguistic border, districts on both sides of this border in the eastern and western parts of Belgium show few signs of integration.

In the first chapter we show that housing markets and prices *within* a single country can move relatively independently because of a different language/culture. In the **second chapter**, we investigate whether there are interactions between housing markets that are located on the opposite side of a *national border*. While the literature on "border effects" in the context of international trade patterns has flourished ever since the seminal contributions of McCallum (1995) and Engel & Rogers (1996) in the *Amer-*

ican Economic Review, there are few studies investigating border effects in housing markets. Due to the fact that mortgage and housing markets are mostly organized on a national level, large differences in housing prices may nonetheless occur along national border. A recent study by Micheli *et al.* (2014), for example, shows that the ask prices of comparable houses drop by about 16% when crossing the Dutch-German border. Despite that we know from administrative data, anecdotal evidence and popular media that housing prices in the Netherlands were generally much higher than in Belgium in recent decades, there are no studies investigating to what extent the higher Dutch housing prices spilled over onto the prices of Belgian properties near the Dutch-Belgian border. Using detailed individual transaction and employing the hedonic pricing method (Rosen, 1974) I examine to what extent "proximity to the border" affects housing prices in Belgium.[2] The results from various (spatial) hedonic models reveal that the prices of Belgian properties closer to the border are, *ceteris paribus*, higher as a result of spatial arbitrage. The estimated arbitrage effect, however, is decreasing over time as a result of the housing price bust in the Netherlands in the aftermath of the global financial crisis of 2007-2008. I also show that the arbitrage effect is not constant across space. Using information concerning the previous of buyers for a sub-sample of all transactions, it is furthermore shown that housing markets are not necessarily efficient as Dutch buyers, on average, pay a premium relative to their Belgian counterparts. The results in this chapter, in some sense, are opposed to those presented in the previous chapter as they show that interactions between housing markets do not necessarily stop at national borders. I believe that this apparent contradiction shows that it is crucial to take into account the nature of a border.

Where the first two chapters of this dissertation focus on spatial spillovers between different housing markets on a regional level, the third chapter zooms in and examines spillover effects at the neighborhood level. Housing economists have acknowledged that housing is a (highly) heterogeneous good and have developed methods, such as the hedonic pricing method (Rosen, 1974), to deal with this inherent heterogeneity. While the view that the housing commodity can characterized as a "*bundle of attributes*" has been generally accepted, at a slightly more aggregate level, neighborhoods can be viewed as "*bundles of properties*". Despite that countless studies have

[2] Due to data limitations, it is unfortunately not possible to calculate the "gap" at the border.

shown that different bundles of housing characteristics are valued differently, there are virtually no studies that investigate whether different bundles of properties in an otherwise equivalent neighborhood are valued differently. In order to fill the void in the literature, in the **third chapter**, we combine a large and detailed dataset of approximately 6,100 dwellings that were sold between 2003 and 2015 in the Flemish part of the Brussels Metropolitan Area with administrative and geospatial data concerning the composition of the housing stock at the neighborhood level. We examine whether the prices of sold properties are affected by the average values and diversity measures of several property characteristics of neighboring properties. In a sense we thus look at spatial spillover effects at the neighborhood level. The results of various (spatial) hedonic models indicate that the price of a property is increasing in the average size and year of construction of neighboring properties, but decreases with the average plot size of neighboring properties. Our findings thus provide support for the *tax capitalization hypothesis* of Hamilton (1976), who argues that smaller dwellings benefit from the presence of larger properties. The results also reveal that home values are higher in neighborhoods that are characterized by low levels of diversity in building types, the year of construction, size, and the shapes of buildings. Our findings furthermore indicate that differences in the composition of the housing stock at the neighborhood level are responsible for price differentials as large as 12%. More generally, the results presented in this paper have potentially far-reaching consequences for real estate professionals, policy makers and urban planners, since they provide some guidelines on how think about the development of new neighborhoods. In my opinion this chapter also provides an interesting example of the use of geospatial data, which is just one example of what economists have dubbed “Big data”, in modern day econometrics.

While the third chapter acknowledges that houses and neighborhoods are heterogeneous, buyers and sellers of real estate are also heterogeneous in a multiplicity of dimensions. Buyers, for example, can have a different willingness to pay for the same property. Buyers and sellers, at the same time, can also be heterogeneous in their search costs. Buyers and sellers with higher search costs have traditionally been more inclined to hire real estate agents, who help facilitating the sales process and serve as intermediaries between buyers and sellers. In return, these real estate agents traditionally charge a service fee. Despite low barriers to entry and housing price movements, these service fees are typically relatively stable and expressed as a percentage of the eventual sales price (e.g. 6% in the US[3]). This had led many people to conclude that the real estate brokerage industry is characterized by a lack of price competition (see e.g. USDOJ, 2007; Anglin & Arnott, 1999). Advances in modern information and communication technologies, however, have put the fixed commission rates under pressure by allowing sellers to sell their property more easily themselves and lowering search costs for both buyers and sellers. Recent studies for the US (e.g. Schnare & Kulick, 2009; Wiley *et al.*, 2012) provide some evidence of price competition in recent years. The obvious question that arises is: "*is more price competition between real estate agents necessarily beneficial from a social point of view?*" In **chapter 4**, which is based on joint work with Bert Willekens, Maarten Goos and Erik Buyst, a theoretical model of imperfect broker competition is developed. The theoretical results show that, under a realistic set of assumptions and heterogeneity in the valuations of buyers and sellers for the services provided by real estate agents, neither the monopoly service fee, nor the service fee that is charged under Bertrand competition is socially efficient. Instead, there exists an inverse u-shaped relationship between the degree of competition and the welfare generated by the brokerage industry. This starkly contrasts the traditional paradigm that more competition is always better. We furthermore show that free broker entry in the presence of fixed operating costs always results in excessive private entry, which is consistent with the findings of Mankiw & Whinston (1986). In the empirical part of this chapter we show that, under simple parametric assumptions, the structural parameters of the model are identified. The model is subsequently calibrated for the Belgian real estate brokerage industry and welfare counterfactuals are performed. The results presented suggest that the observed average commission rate of 4.3% is below the socially commission rate, which is estimated to range between 5.1% and 24%. A welfare gain of 1% to 11% could be established when regulating broker service fees, given the number of brokers that currently operate in the market. When also regulating broker entry, a further welfare gain of 7% to 69% could be realized. Various other policy relevant counterfactuals are constructed and discussed.

So far, we have discussed the unique set of features that underpin the housing commodity and provided a short summary of the four chapters presented in this thesis. All the chapters presented are in some way related to the

[3]See, for example, Hsieh & Moretti (2003).

observations of Quigley (1979, 2002) and Maclennan (1982). As a general context, this thesis was written in the midst of a period of what I believe to be are major changes to the field of (housing) economics. Modern information and communication technologies have both altered the way in which information is gathered and the way and frequency in which it is distributed. The geospatial data used in chapter 3, for example, are just one example of the ever increasing availability of what are known as "Big Data". These new sources of data allow researchers to answer novel and interesting research questions. Housing markets are typically characterized as "thin" markets, which suggests that there might be "bargaining effects". These, however, can only be studied when the appropriate data is available. The results presented in chapters 2 and 4 suggest that bargaining-effects are non-negligible. At the same time, however, big data presents new challenges. The large datasets, for example, require more computational power. Moreover, as shown in the fourth chapter, the developments in ICT are also likely to change existing market structures. This, consequently, raises new research questions and calls for novel methodologies.

Despite that researchers and policy makers can use these new sources of data to address research/policy questions, there are some (spatial) aspects that are not easily quantifiable. Since the housing commodity is inherently spatial, housing economists have also employed methods developed in the domain of spatial econometrics. While the theoretical foundations for models using cross-sectional data were already established in the eighties (e.g. Anselin, 1988) and are by now well-integrated in mainstream economics/econometrics, methods and models using spatial panel data were developed more recently.[4] Authors, such as Pesaran (see e.g. Breitung & Pesaran, 2008; Chudik & Pesaran, 2011; Pesaran, 2004, 2007a, 2007b, 2011; Pesaran & Tosetti, 2011) and Baltagi (2008), have made important contributions in recent years. As housing markets are inherently spatial, we believe that these types of models are particularly eligible for analysing housing prices, both at the micro- and macro level. The first chapter presented in this dissertation provides just one contribution to this growing literature.

[4]Spatial Error Models (SEM) and Spatial Lag Models (SAR), which allow for spatial autocorrelation in the error terms or dependent variable, are for example used in chapters 1, 2 and 3 of this dissertation.

Chapter I

Spatial and Temporal Diffusion of Housing Prices in the Presence of a Linguistic border: Evidence from Belgium

1 Introduction

In this paper we study the effects of the language border in Belgium on the spatial and temporal diffusion patterns of housing prices[1]. A large literature has emerged, especially in the United Kingdom but more recently also in other countries, that investigates regional convergence of and spatial and temporal patterns in housing prices. In the UK, many authors have found that regional house prices are interdependent in the long-run and exhibit a distinct spatial pattern over time, rising first in a cyclical upswing in the south-east and then spreading out over the rest of the country (Meen, 1999). This pattern is often referred to as *the ripple effect hypothesis* and arguably has important implications for the functioning of regional labor markets and the regional distribution of wealth and assets, since housing is one of the most important assets of many households. While statistical evidence for the ripple effect hypothesis has expanded over the past decades and novel econometric models have been introduced to test its validity, Meen

[1]This chapter is published in *Spatial Economic Analysis*, 2016, 11(1), 92-122. We are extremely grateful to Geoffrey Meen, Annelore Van Hecke, Frank Verboven, Frank Vastmans, Sven Damen, Geert Goeyvaerts and the participants of the housing economics workshops of the European Network for Housing Research conferences in 2012 (Lillehammer, Norway) and in 2013 (Tarragona, Spain) for valuable comments and suggestions on earlier drafts of this work.

(1999) notes that there are fewer studies that provide convincing economic explanations. He argues that the observed ripple effect may be the result of (1) migration, (2) equity transfer, (3) spatial arbitrage, and/or (4) spatial patterns in the determinants of housing prices. It is intuitive that especially the first three of these mechanisms might be (strongly) influenced by the presence of a (language) border. Descriptive results presented in table 3 suggest that the degree of coherence among regional housing markets is indeed higher within each linguistic region. This indicates that the language border may play an important role.

Using mix-adjusted house price transaction data for 20 *a priori* defined districts in Belgium and an extended version of an econometric model that was recently proposed by Holly *et al.* (2011) to cope with the unique federal structure of Belgium, we provide evidence for existence of the ripple effect hypothesis and show that the linguistic border plays an ambiguous role. Our findings show that an initial shock to house prices in the dominant region (Antwerp) is quickly absorbed by districts located along the north-south axis of Belgium, which constitutes Belgians economic spine, and thus crosses the language border. We furthermore show that housing prices in the peripheral regions in the eastern and western part of the country converge solely with respect to their neighbors *within* their respective linguistic region. These results suggest that housing markets in the peripherally located districts in the eastern and western parts of Belgium are far less integrated with their neighbors across the linguistic border than districts centrally located districts along the north-south spine.

We contribute to the existing literature in a number of ways. Despite that the literature on spatial and temporal diffusion patterns of regional housing prices has also expanded outside the UK, the current paper is the first study that provides evidence for the existence of the ripple effect hypothesis for Belgium. Secondly, the current study contributes to the literature on the existence of and the mechanisms that lay behind border effects. Similar to the literature on the ripple effect hypothesis there is a growing body of literature that provides statistical evidence for the existence of border effects, but there are fewer studies that investigate the reasons behind this. In the current study we aim to fill this void. Using the unique federal structure of Belgium as a case-study, we show that linguistic differences influence spatial and temporal diffusion patterns of housing prices. Specifically for Belgium we contribute to the existing literature by showing that the linguistic border does not only matter in the short-run, but that the observed effects are also persistent in the long-run.

Despite that a similar study as the current one might only be carried out in a few other multilingual countries (e.g. Switzerland and Canada), the results presented here shed some light on why border effects are still observed as, for example, reported by Chen (2004). Similar to a study by Ferreira-Lopes & Sequiera (2012) our results might also hold lessons for the EU given the many linguistic and cultural differences among its member states.

The rest of the paper is set out as follows. In section 2 we provide an overview of the existing literature that is related to our study. In section 3, subsequently we present the data and provide a brief introduction into the federal and linguistic structure of Belgium. In section 4 we present the methodology used in the subsequent empirical analysis. The estimation results and the simulated Generalized spatio-temporal Impulse Response Functions are presented and discussed in section 5. Finally, section 6 concludes.

2 Literature review

Spatial and temporal diffusion patterns of regional housing prices have been studied in the UK and other countries (e.g. Stevenson (2004) for Ireland, van Dijk *et al.* (2010) for the Netherlands, Kuethe & Pede (2011) for the US). In the UK many researchers (e.g. MacDonald & Taylor (1993), Alexander & Barrow (1994), Holmes & Grimes (2008), and Abbott & De Vita (2013)) have focused on long-run relationships between regional housing prices. These and others authors (e.g. Giussani & Hadjimatheou (1991), Ashworth & Parker (1997) and Meen (1999)) have also investigated (short-run) causality between regional house prices, which has often been referred to as the ripple effect hypothesis. As Meen (1999) points out, both strands of literature are closely related since the ripple effect hypothesis implies that "*short-term variations in regional price differentials can be very large indeed, but in the longer term some normal relative price pattern tends to be restored.*" Despite the large interest and statistical evidence for the ripple effect hypothesis in the UK, Meen (1999) argues that are few studies providing convincing economic explanations. He puts forward four possible explanations, namely (1) migration, (2) equity transfer, (3) spatial arbitrage, and

(4) spatial determinants of house prices. A fifth possible explanations that is examined in Meen & Andrew (1998) are leads and lags in house prices. It is intuitive that a (language) border might have important implications for the relative strength of these mechanisms. Although we do not explicitly model the aforementioned explanations in our econometric framework we provide some insights into the mechanisms that are responsible for the observed ripple effect in Belgium.

Given that we study the effects of a (language) border on the spatial and temporal convergence and diffusion of housing prices, our paper also relates to the literature on border effects. The effects of (international) borders have mostly been studied in the context of international trade patterns where many authors have observed that trade volumes are much higher between regions within a single country than equidistant regions that are located in different countries. In a seminal paper by McCallum (1995), for example, the author find that the relatively innocuous Canada-US border continues to play a decisive effect on trade patterns. In another seminal paper by Engel & Rogers (1996) the authors examine deviations from the law of one price using CPI data for US and Canadian cities for 14 categories of consumer products. Their results indicate that while the distance between cities explains a substantial amount of the variation in the data for similar goods, the price variation is much higher for two cities located on opposite sides of the border than for two equidistant cities in the same country. In the context of regional housing prices the effects of a national border have been studied by Stevenson (2004). The author investigates house prices diffusion within the Republic of Ireland and between the Republic and Northern Ireland and finds that the Northern Irish market is more linked with that of the Republic than with the rest of the UK. All these papers suggest that national borders (continue to) matter. Chen (2004) also investigates why national borders continue to matter in the EU and finds that, contrary to previous findings in the literature, trade barriers do provide an explanation. In particular, she finds that technical barriers to trade, together with product-specific information costs, increase border effects. Although the author does not specify the nature of these information costs, it is intuitive that costs related to linguistic differences may be one of them. In the current study we examine the effect of the linguistic border that divides Belgium into two distinct regions on the spatial and temporal propagation of housing prices. While many studies have estimated the effects of national borders, we are one of the first studies examining the effects of a well-defined border within a single country. Given the almost identical institutional set-up on both sides of the language border we can safely rule out this potential explanation for the border effects that have been observed in previous studies.

A third strand of literature that relates to the current study focuses on the economics of language. While the economics of language has especially been studied in the context of labor markets (e.g. Chiswick, 2008) and the form of human capital they provide to immigrants and native-born linguistic minorities, there have been a few other studies that investigate the effects of language (differences) on economic outcomes. Schulze & Wolf (2009) show in a paper titled "*On the Origins of Border Effects: Insights from the Habsburg Empire*" that borders continue to matter in periods of increasing economic integration. They furthermore show that ethno-linguistic networks had persistent trade diverting effects in the multi-national Habsburg Empire prior to the First World War. More specifically, their results indicate that political borders became visible in the economy from the mid-1880s onwards. They attribute this result, which they call "border before a border", largely to the ethno-linguistic composition of the population across the different regions. In a more recent study Ferreira-Lopes & Sequeira (2012) investigate the degree of interdependence among business cycles in different regions in modern-day Switzerland. They find that the cantons - a level of aggregation that is similar to that used in the current study - are closely related, but there are dynamic effects toward more "independent" business cycles.

The current paper also relates to two studies that examine the effects of the linguistic border in Belgium on (regional) housing markets. Goffette-Nagot *et al.* (2011) find that the linguistic border acts as a strong barrier in the spatial pattern of land prices. De Bruyne & Van Hove (2013) also find important differences between the different Belgian linguistic regions using municipality-level data on housing prices. While both of the aforementioned studies use cross-sectional data on land and housing prices, in this paper, we study the effects of the language border on the long-run convergence and spatial and temporal diffusion patterns of housing prices by using a large panel dataset of regional housing prices.

3 Data and the linguistic border

Unlike many studies in the UK and the US, we unfortunately do not have mix-adjusted volume-weighted hedonic price indices at our disposal. All data concerning housing prices in Belgium are recorded by *Statistics Belgium* and are published at a quarterly frequency for all municipalities and higher levels of aggregation. The published data comprise the number of transactions and the average, median and other quantile prices of all transactions that occurred in a particular municipality in a particular period. These series are reported for four different categories, namely (1) apartments, (2) "ordinary" dwellings, (3) villa's, mansions and bungalows, and (4) land, where we ignore the latter two.[2] We control for changes in composition and location using the following simple weighting procedure:

$$p_{i,t} = \Sigma_{k=1}^{K} \bar{w}_{i,k} p_{i,k,t} \text{ where } \bar{w}_{i,k} = T^{-1} \Sigma_{t=1}^{T} w_{i,k,t} \tag{I.1}$$

Where subscript i denotes the district, k the respective category and t the quarter of sale. Since we want to study within country diffusion patterns and our data is available at the municipal level, we exclude data from transactions that occurred in municipalities located along borders with neighboring countries and also exclude data from coastal municipalities. While house prices in the border regions, especially along the border with the Netherlands, are likely to be influenced by the inflow of (fiscal) migrants, the Belgian coastal municipalities are characterized by a high share of (transactions of) secondary homes. After exclusion of these transactions we aggregate the data at the level of the *judicial* districts[3], since these largely correspond with economic entities - i.e. city and its respective agglomeration and rural area. Another reason for the aggregation procedure used it that the number of transactions at the municipal level is very often too limited. A map of the different districts used in the analysis is presented in figure 1.

[2] Villa's, mansions and bungalows is a small subcategory with in general high transaction prices. Inclusion of these transactions might lead to unwanted outliers. Therefore these transactions are dropped out of our sample. (Residential) land is left out for obvious reasons.

[3] Before the reform of 2012/2013, there were 27 judicial districts in Belgium. In the empirical analysis we use 20 districts. To this end we have merged the districts of Tournai and Mons (Tournai), Verviers and Eupen (Verviers), Arlon, Neufchâteau and Marche-en-Famenne (Arlon), Huy and Liège (Liège), and Tongeren and Hasselt (Hasselt).

Figure 1: *Regions, districts, coastal municipalities and border municipalities*

Source maps: Belgian HISGIS

Once we have constructed the regional housing price series, we employ the national Consumer Price Index (CPI)[4] to deflate nominal housing prices in line with the existing literature. Finally, we collected data from the National Bank of Belgium concerning the evolution of real GDP in Belgium, which are available at a quarterly frequency from 1980Q1 up to 2011Q1 (T=125). An overview of the raw data is presented in appendix A.

Now that we have discussed the data used in the empirical analysis, it is necessary to provide some insights on the origins and the nature of the linguistic border in Belgium which spans over approximately 200 kilometers through Belgium and is the official divide between the Northern Dutch-speaking Flemish Region and the southern French-speaking Walloon Region. The linguistic border in Belgium is part of a much larger linguistic (and cultural) border that divides western Europe in two large areas. With the notable exceptions of Belgium and Switzerland, this continental division coincides largely with national borders. South of the divide are countries, such as France, Italy and Spain, where Romanic languages are spoken. In

[4]Data concerning the CPI is provided by Statistics Belgium and is unfortunately not available on a more disaggregated level.

countries that are located north of this border, such as the Netherlands, Germany and Austria, Germanic languages are spoken. While the linguistic conflict in Belgium was mostly non-violent, many wars have been fought between countries on both sides of this continental divide in Europe over the past centuries (e.g. the Franco-Prussian War in 1870-1871). The linguistic border in Belgium, which was explicitly fixed in Belgian law in 1962/1963 after an extensive period of linguistic conflict, implies among other things that all official documents are only to be drafted in the official language spoken in the respective Region. An exception however is the Brussels Capital Region, which is an officially bilingual enclave located in the Flemish Region.

Now that we have discussed the origins, it is necessary to provide some further/deeper insights. Up to 1947, as a part of the decennial census, the government counted the use of the different languages in every municipality. Although these data date back almost 70 years, they shed some light on the "hardness" of the border in daily life. Data from this last language count in 1947 suggest that approximately 94.7% and 95.4% of all people in respectively the Flemish Region and the Walloon Region mostly or exclusively spoke the official language in their region, which suggests that there were two distinct linguistic regions in Belgium. Since no official language counts were held after 1947, we cannot evaluate whether this pattern remains consistent over time using data from language counts. More recent data concerning migration patterns at the provincial level presented in table 1, however, show that migration is much lower between provinces on opposite sides of the linguistic border than between provinces in the same linguistic region.

Table 1: *Overview migration patterns within and between linguistic regions*

Origin/destination	Within	Between	BCR
Antwerp	86.7	4.7	8.5
Flemish Brabant	50.4	18.1	31.4
Walloon Brabant	53.3	13.2	33.3
West Flanders	76.6	15.4	7.9
East Flanders	83.9	7.7	8.3
Hainaut	58.6	18.2	23.1
Liège	62.2	17.5	20.1
Limburg	62.2	9.1	5.4
Luxembourg	81.5	6.3	12
Namur	81.9	4.8	13.2
Average	72	11.5	16.3

Note: the migration data, which are for the year 2011, were collected from the website of Statistics Belgium.

One might argue, however, that this higher probability of moving *within* each linguistic region is due to other reasons, such as proximity. The results of a simple regression analysis, presented in table 2, where we control for the distance between different provinces (common border) indeed confirm that mutual distance explains a substantial amount of the variation in migration patterns. The results, however, also reveal that the language border continues to play an important role, since households are much more likely to move *within* their respective linguistic region.

Table 2: *Regression analysis migration patterns*

Variable	$\hat{\beta}$	$\hat{\sigma}$
Same linguistic region	0.093***	0.016
Common border	0.118***	0.016
Brussels CR	0.141***	0.025
Constant	0.01	0.01
R-sq.	0.581	
Obs.	110	

Note: the dependent variable is the percentage of people migrating from province j to province i. ***, ** and * signify that the test rejects the null at respectively the 1, 5 and/or 10% level.

So far, we have established that Flanders and Wallonia are two distinct regions where most people mostly or exclusively speak the official language of their respective region on a daily basis. We have furthermore shown that both regions are characterized by a low rate of interregional migration.

4 Methodology

We continue in this section by proposing our empirical framework. In section 4.1 we discuss the spatial weights matrices used in the empirical analysis and in section 4.2 we propose the econometric model.

4.1 Spatial weights matrices

Firstly, it is important to introduce the mathematical representation of the spatial structure in the econometric model. Spatial structure has very often been operationalized by so-called spatial weights matrices, which are *a priori* defined by the researcher. Given that there is no definite answer how space should be mathematically represented and the estimation results are conditional upon the mathematical structure of space, it is important to investigate whether the results are driven by the chosen weighting scheme. In the current study we use 2 of these schemes to ensure the robustness of our results with respect to our choice concerning the mathematical representation of spatial structure in the data. The first scheme we employ in the current study is the so-called *contiguity criterion*, which has been used on numerous occasions in the literature. Districts i and j are considered to be *neighbors* whenever they share a common border.

$$W_{i,j}^{\text{con}} = \begin{cases} 1 \text{ if } i \text{ and } j \text{ are neighbors} \\ 0 \text{ otherwise} \end{cases} \tag{I.2}$$

A second weighting scheme that frequently appeared in the spatial econometrics literature is the *inverse distance* scheme which relates to Waldo Tobler's (1979) *first law of geography*[5].

[5] "*Everything is related to everything else, but near things are more related than distant things.*"

$$W_{i,j}^{\text{id}} = \frac{1}{d_{i,j}} \tag{I.3}$$

Where $d_{i,j}$ denotes the distance between districts i and j and is measured as the crowfly distance between the major cities of each pair of districts i and j.

Since we especially want to focus on the effects of the linguistic border in the spatial and temporal propagation of housing prices, we have to take into account in which linguistic region districts i and j are located. We therefore split the spatial weights matrices presented into four separate spatial weights matrices by pre-multiplying the spatial weights matrices (we use the Hadamard-product, i.e. element-by-element) $W_{i,j}^{\text{con}}$ and $W_{i,j}^{\text{id}}$ with $W_{i,j}^{\text{o}}$ and $W_{i,j}^{\text{c}}$, the elements of which are equal to 1 whenever the districts are located in the same (other) linguistic region and equal to 0, otherwise. We thus construct the following spatial weights matrices:

$$W_{i,j,o}^{\text{con}} = W_{i,j}^{o} \circ W_{i,j}^{\text{con}} \tag{I.4}$$
$$W_{i,j,c}^{\text{con}} = W_{i,j}^{c} \circ W_{i,j}^{\text{con}} \tag{I.5}$$
$$W_{i,j,o}^{\text{id}} = W_{i,j}^{o} \circ W_{i,j}^{\text{id}} \tag{I.6}$$
$$W_{i,j,c}^{\text{id}} = W_{i,j}^{c} \circ W_{i,j}^{\text{id}} \tag{I.7}$$

Finally, as conventional in the spatial econometrics literature, the resulting four spatial weights matrices ($W_{i,j,o}^{\text{con}}$, $W_{i,j,c}^{\text{con}}$, $W_{i,j,o}^{\text{id}}$, and $W_{i,j,c}^{\text{id}}$) are row-normalized.[6]

4.2 Econometric model

The econometric model used in the subsequent empirical analysis is a spatio-temporal model for housing prices that is similar to that proposed by Holly *et al.* (2011) and has also been applied to the Chinese housing market by Gong *et al.* (2014). As in their contributions, we are interested in the propagation of (log) housing prices, $p_{i,t}$, over time (indexed by $t = 1, 2, ..., T$) and space (indexed by $i = 1, 2, ..., N$). We furthermore want to allow for the possibility of a dominant region, region 0, and error correcting mechanisms. Whilst shocks to the dominant region are propagated to other regions immediately

[6]This transformation assures that every districts has the same number of neighbors.

shocks to the remaining regions have no immediate impact on the dominant region. The error correcting mechanisms take into account possible long-run equilibria among housing prices in different regions and are allowed for when the co-trending vector is found to be a valid representation of the data. For the dominant region, region 0, the following house price equation thus applies:

$$\begin{aligned}\Delta p_{0,t} = \phi_{0,s_o}(p_{0,t-1} - \bar{p}^{s_o}_{0,t-1}) + \phi_{0,s_c}(p_{0,t-1} - \bar{p}^{s_c}_{0,t-1}) + \alpha_0 \\ +\beta_{0,1}\Delta p_{0,t-1} + \gamma_{0,1}\Delta\bar{p}^{s_o}_{0,t-1} + \delta_{0,1}\Delta\bar{p}^{s_c}_{0,t-1} + \epsilon_{0,t}\end{aligned} \tag{I.8}$$

where $\Delta p_{o,t}$ denotes the growth rate of housing prices in the dominant region at time t, $\Delta\bar{p}^{s_o}_{0,t-1}$ represents the growth rate of housing prices in neighboring regions on the same side of the linguistic border at time $t-1$, and $\Delta\bar{p}^{s_c}_{0,t-1}$ the growth rate of housing prices in neighboring regions on the opposite side of the linguistic border at time $t-1$. Shocks to housing prices in the previous period in neighboring regions are thus allowed to have an effect on the growth rate of housing prices in the current period in the the dominant region. $p_{0,t-1}$, $\bar{p}^{s_o}_{0,t-1}$ and $\bar{p}^{s_c}_{0,t-1}$, respectively, denote the natural logarithm of real housing prices at time $t-1$ in the dominant region, neighboring regions on the same side of the linguistic border, and neighboring regions on the opposite side of the linguistic border. The growth rate of (real) housing prices in the dominant region can thus be affected by deviations from long-run equilibria and the growth rates of housing prices in both the own and neighboring regions in the previous period.

For the remaining regions the following price equation is estimated:

$$\begin{aligned}\Delta p_{i,t} = \phi_{i,0}(p_{i,t-1} - p_{0,t-1}) + \phi_{i,s_o}(p_{i,t-1} - \bar{p}^{s_o}_{i,t-1}) + \phi_{i,s_c}(p_{i,t-1} - \bar{p}^{s_c}_{i,t-1}) \\ +\alpha_i + \beta_{i,1}\Delta p_{i,t-1} + \gamma_{i,1}\Delta\bar{p}^{s_o}_{i,t-1} + \delta_{i,1}\Delta\bar{p}^{s_c}_{i,t-1} + \kappa_i\Delta p_{i,t} + \epsilon_{i,t}\end{aligned} \tag{I.9}$$

Where $p_{i,t}$ denotes the (natural) logarithm of real housing prices in district i at time t, and $\Delta p_{i,t}$ its growth rate. $\bar{p}^{s_o}_{i,t}$ denotes the (natural) logarithm of real housing prices in neighboring regions on the *same* side of the linguistic border and $\bar{p}^{s_c}_{i,t}$ denotes the equivalent for neighboring regions on the *opposite* side of the linguistic border. Similarly, $\Delta\bar{p}^{s_o}_{i,t-1}$ and $\Delta\bar{p}^{s_c}_{i,t-1}$ denote the spatially and temporally lagged growth rates of housing prices for districts on respectively the same and the opposite side of the language border. Note that the model presented here is a 'pure' price diffusion model where prices in every region i solely react to house price developments in the dominant and neighboring regions. Since it is likely that house prices changes are the result of (macro)economic developments we also include the change in the natural logarithm of real GDP at time t, ΔGDP_t, in our regression analysis for every region i. Furthermore, note that the model presented here is a first-order linear representation. In the empirical application, however, we allow for higher order lags. Due to the presence of $\Delta p_{0,t}$ in the price equations for the remaining regions we have to perform, similar to Holly *et al.* (2011), Wu-Hausman type of exogeneity tests (Wu, 1973). Once we have estimated the model, we can also simulate the appropriate Generalized spatio-temporal Impulse Response Functions.

5 Results

5.1 Long-run equilibria

Before we present the results of the selection procedure used to determine a suitable candidate for the dominant region and the results from our econometric model, we first provide some insights into the degree of integration among housing prices in the different Belgian districts. To this end, we use the recently developed pair-wise approach (Pesaran, 2007) that was also used in an earlier study by Abbott & De Vita (2013). We perform unit root tests on all $N(N-1)/2$ pairs of house price differentials $(p_{i,t} - p_{j,t})$. Descriptive statistics concerning the fraction of pairs for which the house price differential is stationary at the 5% level using standard ADF-tests are presented in table 3.

Table 3: *Fraction of districts that are cointegrated at the 5% level using log real house prices (1973Q1-2011Q3)*

All	Within linguistic Region			Across border
	Total	Flanders	Wallonia	
0.468 (190)	0.563 (94)	0.454 (66)	0.821 (28)	0.375 (96)

Note: the number of observations is displayed between brackets. Cointegration statistics are calculated using pairwise ADF-tests on regional house price differentials where the optimal lag is calculated using the SIC-criterion. A full table of all cointegration statistics is available upon request from the authors.

The results generally indicate that there is a high degree of interdependency among the different Belgian districts, given that 46.8% of the 190 pairs of house price differentials are stationary at the 5% level. Since Belgium is only a small country this is intuitive. The results furthermore reveal that the fraction of regional house price differentials that is stationary is higher *within* each linguistic area (56.3%), which indicates that either language, distance or other factors play a role. The results finally reveal that the degree of coherence is higher among Walloon districts than among their Flemish counterparts.

5.2 Choice of the dominant region

In the empirical model described in section 4.2 we allow for a dominant region, where shocks to this dominant region are contemporaneously and spatially propagated to the remaining regions without immediate feedback effects. This type of model can be characterized as a VAR model with a dominant unit (Chudik & Pesaran, 2011) and has previously been applied by Holly *et al.* (2011) for the UK and Gong *et al.* for the Chinese housing market. In the current study we estimate the following bivariate VAR(4) models with error correcting coefficients, which allows us to assess whether housing prices in a certain region are long-run forcing, in the sense of Granger & Lin (1995), upon prices in the remaining regions:

$$\begin{aligned}
\Delta p_{i,t} &= \phi_{i,j}(p_{i,t-1} - p_{j,t-1}) + \Sigma_{l=1}^{4} a_{i,j,l}\Delta p_{i,t-l} + \Sigma_{l=1}^{4} b_{i,j,l}\Delta p_{j,t-l} + \epsilon_{i,j,t} \\
\Delta p_{j,t} &= \phi_{j,i}(p_{j,t-1} - p_{i,t-1}) + \Sigma_{l=1}^{4} a_{j,i,l}\Delta p_{j,t-l} + \Sigma_{l=1}^{4} b_{j,i,l}\Delta p_{i,t-l} + \epsilon_{j,i,t}
\end{aligned} \tag{I.10}$$

Where the error correction coefficients, $\phi_{i,j}$ and $\phi_{j,i}$, and their associated t-ratios are estimated using a SUR-algorithm[7]. The results are presented in table 4.[8]

[7] Seemingly Unrelated Regressions

[8] The significance levels for all the error correction coefficients are presented in table B.1 in appendix B.

Table 4: *Error correction coefficients in cointegrating bivariate VAR(4) of log real house prices of Antwerp and other Belgian districts (1973Q1-2011Q3)*

District	$\hat{\phi}_{0,i}$	$t_{0,i}$	$R^2_{0,i}$	$\hat{\phi}_{i,0}$	$t_{i,0}$	$R^2_{i,0}$
Mechelen	0.016	0.337	0.211	-0.167***	-2.923	0.304
Turnhout	0.043	1.305	0.159	-0.140***	-3.164	0.245
Brussels	-0.054	-0.933	0.273	-0.118**	-2.056	0.241
Leuven	-0.093*	-1.756	0.254	-0.231***	-3.292	0.264
Nivelles	0.042	0.767	0.23	-0.168***	-2.781	0.235
Bruges	-0.017	-0.398	0.2	-0.115**	-2.348	0.294
Kortrijk	0.003	0.09	0.189	-0.091**	-2.421	0.297
Veurne	0.008	0.268	0.165	-0.189***	-3.627	0.297
Dendermonde	0.014	0.291	0.173	-0.170***	-3.084	0.288
Ghent	-0.018	-0.588	0.182	-0.014	-0.394	0.322
Oudenaarde	0.029	0.916	0.194	-0.143***	-3	0.326
Charleroi	0.033	1.095	0.177	-0.052*	-1.871	0.313
Tournai	0.021	0.563	0.194	-0.095**	-2.477	0.317
Liège	0.024	0.737	0.215	-0.075**	-2.413	0.321
Verviers	0.034	0.957	0.256	-0.168***	-2.918	0.296
Hasselt	0.017	0.555	0.153	-0.103***	-2.598	0.313
Arlon	0.068*	1.902	0.197	-0.209***	-3.375	0.276
Dinant	0.03	0.86	0.151	-0.140***	-2.867	0.26
Namur	0.018	0.421	0.195	-0.090*	-1.884	0.333

Note: the table displays the results of the pair-wise long-run causality test. ***, ** and * indicate that the error correction coefficient is significant at the 1, 5 and/or 10% level. A full table for all the districts is available from the authors upon request.

The estimates presented in table 4 indicate that the district of Antwerp is the most suitable dominant region, given that housing prices in the district of Antwerp are long-run causal (Granger & Lin, 1995) upon prices in all other regions, except for the district of Ghent. The results presented in table B.1 in appendix B furthermore confirm that Antwerp is the most suitable candidate for the dominant region. This result might seem counter-intuitive since Brussels is the largest city in Belgium, centrally located and home to many national and international organizations. Our intuition however is that Antwerp, with its port, is a more suitable candidate as it is more prone to (international) economic shocks. The port of Antwerp is Europe's second largest port with a large hinterland including Western Germany and Northern France (Loyen *et al.*, 2003) and hosts one of the world's most important clusters of chemical industry.[9] Data from the National Social Security Office

[9] The port of Antwerp provided employment for 145,836 full time equivalents (FTE)

in Belgium furthermore reveal that while 38.5% of the working population in the Brussels Capital Region is employed by the public sector, this sector is responsible for only 22% of total employment in Antwerp, which suggests that economic shocks might be felt - and capitalized into housing prices - sooner in the port town. Given the results presented in this subsection we start from the hypothesis that Antwerp is the dominant region in Belgium and estimate the full model accordingly.

5.3 Estimation results

So far, we have presented results which indicate that housing prices in districts within each linguistic area display a higher degree of interdependence and the district of Antwerp is likely a suitable candidate for the dominant region. Obviously, these simple analyses do not capture all possible interdependencies between housing prices in different districts. Therefore, we estimate the full model presented in section 4.2 where Antwerp acts as the dominant region and spatial relationships are split up into two separate components (within and across linguistic region) using the contiguity criterion discussed in section 4.1 to construct the appropriate spatial weights matrices. All price equations are estimated using OLS and lag orders are selected using the SIC criterion with a maximum lag order of 4. The results are presented in table 5.

A first glimpse at the estimates presented in table 5 reveals that housing prices in Belgium are highly integrated among each other, both in the short run and in the long run. The table also shows that districts react heterogeneously with respect to shocks in (real) GDP growth. We furthermore observe that the effect of shocks to Antwerp, are contemporaneously and spatially propagated to 14 out of the 19 remaining districts, which suggests that Antwerp is indeed a suitable candidate for the dominant region. The Wu-Hausman test-statistics that are presented in the second to last column, also suggest that Antwerp is an eligible candidate given that the null-hypothesis of exogeneity is rejected for only one (Nivelles) of the 19 districts. The test-statistics thus ensure that the results are not subject to

in 2010, of which 60,509 direct and 85,327 indirect. The 10 largest employers were: BASF, BNRC Group, Public sector, Antwerp Port Authority, General Motors Belgium, ExxonMobil Petroleum & Chemical, PSA Antwerp, M.S.C. Home terminal, Electrabel and Total Refinery Belgium. The port of Antwerp was responsible for approximately 19.2 billion euros of value added, which was approximately 5.5% of the total value added in Belgium in 2010 (Mathys, 2012).

Table 5: *Estimation results of district specific house price diffusion equation with Antwerp as the dominant region and using the contiguity criterion to construct the appropriate spatial weights matrices (1980Q1-2011Q3)*

									Lag order:				
District	$\hat{\phi}_{i,0}$	$\hat{\phi}_{i,s_o}$	$\hat{\phi}_{i,s_c}$	$\Delta p_{i,t-1}$	$\Delta\bar{p}^{so}_{i,t-1}$	$\Delta\bar{p}^{sc}_{i,t-1}$	$\Delta p_{0,t}$	$\Delta \mathrm{GDP}_t$	$\Delta p_{i,t-1}$	$\Delta\bar{p}^{so}_{i,t-1}$	$\Delta\bar{p}^{so}_{i,t-1}$	WH	R^2_i
Antwerp				-0.29***	0.340***			0.924***	1	1			0.326
				(0.094)	(0.107)			(0.347)					
Mechelen	-0.19***			-0.32***	0.356***		0.317***	-0.53	1	4		-0.787	0.513
	(0.061)			(0.095)	(0.136)		(0.116)	(0.373)					
Turnhout		-0.19***		-0.25***	0.319**		0.355***	0.571	2	1		0.949	0.468
		(0.051)		(0.087)	(0.127)		(0.097)	(0.416)					
Brussels	-0.15***			-0.1	0.034	0.121	0.311***	0.467**	4	1	1	0.856	0.598
	(0.046)			(0.082)	(0.081)	(0.076)	(0.076)	(0.237)					
Leuven	-0.21***			-0.40***	0.032	0.128	0.305***	-0.04	3	4	1	0.096	0.517
	(0.072)			(0.102)	(0.172)	(0.159)	(0.106)	(0.378)					
Nivelles	-0.17***			-0.33***	0.199	0.292***	0.211***	1.07***	1	1	1	-2.095**	0.515
	(0.048)			(0.076)	(0.138)	(0.092)	(0.077)	(0.335)					
Bruges	-0.17***	-0.20**		-0.20**	-0.02		0.422***	1.05***	1	1		0.924	0.521
	(0.048)	(0.085)		(0.092)	(0.104)		(0.095)	(0.386)					
Kortrijk	-0.14***		-0.18*	-0.13	-0.07	-0.06	0.206**	0.565	1	1	1	1.861*	0.345
	(0.032)		(0.099)	(0.114)	(0.115)	(0.107)	(0.098)	(0.383)					
Veurne		-0.71***		-0.03	-0.37**		0.007	0.427	1	1		0.982	0.387
		(0.137)		(0.096)	(0.178)		(0.142)	(0.502)					
Dendermonde		-0.36***		-0.03	-0.18		0.184**	0.621	1	1		1.177	0.386
		(0.084)		(0.107)	(0.12)		(0.08)	(0.427)					
Ghent				-0.59***	0.564***		0.136	0.616	2	2		0.924	0.345
				(0.11)	(0.14)		(0.16)	(0.441)					
Oudenaarde		-0.39***		-0.28***	-0.1	-0.05	0.084	-0.16	1	1	1	0.103	0.399
		(0.084)		(0.091)	(0.222)	(0.142)	(0.148)	(0.541)					
Charleroi		-0.34***		-0.15	-0.19		0.188***	0.307	1	1		0.949	0.544
		(0.054)		(0.094)	(0.121)		(0.069)	(0.269)					
Tournai		-0.23**		-0.37***	0.302*	0.001	0.252***	0.168	2	4	1	-1.256	0.595
		(0.095)		(0.121)	(0.162)	(0.107)	(0.085)	(0.37)					
Liège		-0.12*	-0.20***	-0.25**	0.087	-0.01	-0.04	0.621*	1	1	1	-0.435	0.576
		(0.076)	(0.06)	(0.104)	(0.109)	(0.08)	(0.079)	(0.326)					
Verviers	-0.17***			-0.36***	0.214*		0.314**	0.449	2	1		-0.451	0.472
	(0.046)			(0.093)	(0.119)		(0.13)	(0.571)					
Hasselt		-0.32***		-0.32***	-0.1	-0.04	0.301***	0.259	1	1	1	0.165	0.441
		(0.088)		(0.118)	(0.108)	(0.097)	(0.105)	(0.603)					
Arlon	-0.20***	-0.29**		-0.15	0.124		0.327***	0.943	2	1		-0.316	0.425
	(0.058)	(0.132)		(0.128)	(0.178)		(0.126)	(0.705)					
Dinant	-0.11**	-0.33**		-0.23**	0.298		0.01	1.02**	1	1		-0.565	0.425
	(0.052)	(0.137)		(0.104)	(0.186)		(0.122)	(0.472)					
Namur	-0.15***			-0.43***	0.288**		0.243***	0.689*	1	1		0.602	0.481
	(0.034)			(0.09	(0.139)		(0.092)	(0.379)					

Note: Standard errors are shown in parentheses. ***, ** and * signifies that the test rejects the null at respectively the 1, 5 and/or 10% level. All regressions include an intercept term, 3 seasonal dummies and an additional dummy variable to account for the change in the classification process that was implemented by Statistics Belgium at 2005Q1.

simultaneity bias.

While the previous paragraph provides a general description of the results presented in table 5, our goal is to study the effects of the linguistic border in the spatial and temporal propagation of housing prices, which necessitates a more thorough look at the results presented above. We already mentioned that the cointegration statistics presented in table 3 and the error correction coefficients presented in table 5 suggest that housing prices in Belgium are highly integrated, where the results presented in table 3 suggested that housing prices are more highly integrated within each linguistic region. The error correction coefficients presented in table 5 show that all districts, except for Ghent, converge with respect to Antwerp, neighboring districts, or both. A more thorough investigation shows that the districts can be split up into approximately 3 different groups. A first group of districts that comprises Mechelen, Brussels, Leuven, Nivelles, Namur and Verviers are districts that are located along, except for Verviers, the economically important north-south axis and converge with respect to the dominant region, Antwerp. Notice that Nivelles, Namur and Verviers are located in the Walloon region, while the remaining districts are Dutch-speaking, which implies that convergence is not strictly limited to districts within the same linguistic region. A second group that comprises Bruges, Arlon, Kortrijk and Dinant are (peripherally located) districts that converge both with respect to the dominant region and neighboring regions. Observe that the error correction mechanism with respect to neighboring districts is larger for all four districts than the error correction mechanism with respect to the dominant region. Furthermore note that although Kortrijk converges with respect to neighboring districts on the opposite side of the linguistic border, the error correction coefficient is only significant at the 10% level. A final group of (peripheral) districts that comprises Turnhout, Dendermonde, Veurne, Oudenaarde, Charleroi, Tournai, Liège and Hasselt are districts that converge solely with respect to neighboring districts. While the districts of Oudenaarde, Tournai, Liège and Hasselt can potentially display convergence with respect to neighboring districts across the linguistic border only the error correction coefficient for Liège is statistically significant.

Our econometric specification does not only allow for long-run convergence, but also allows for short-run dynamics. These are captured by the lagged own price changes and the lagged price changes of neighboring districts. The results point out that 14 out of the 20 districts have negative coefficients for lagged own price changes. Although one might worry about potential seasonality effects, we have already controlled for these effects by adding seasonal dummies in our regressions. To capture spatial spillovers and account for the presence of the linguistic border, we included lagged prices changes of both neighboring districts within the same linguistic regions and neighboring districts across the linguistic border. The estimates display a strong linguistic pattern, where lagged price changes in neighboring districts within the same linguistic region are generally positive (13 out of 20 districts) and statistically significant (8 out of 20 districts), while lagged price changes of neighboring districts across the linguistic border insignificant in 7 out of 8 cases. The notable exception here is the Walloon district of Nivelles, which neighbors the Flemish districts of Leuven and Brussels. These three districts together up to 1995 constituted the province of Brabant and are economically highly integrated due the presence of the (bilingual) Brussels Capital Region (note that the Brussels Capital Region does not coincide with the judicial district of Brussels). Many people working in the Brussels Capital Region commute on a daily basis from municipalities that are located in both Walloon and Flemish Brabant.

5.4 Robustness

Although the model presented previously allows for shocks in (real) GDP to be absorbed in housing prices instantaneously (real) GDP is only available from 1980Q1 onwards, while housing prices are available from 1973Q1. Therefore, we have also estimated the same model without (real) GDP growth (full sample, i.e. $T = 155$). The results, which are presented in table 6, indicate that the degree of interdependence among regional housing prices is even higher as all regions have one or multiple (spatial) error correction terms that are negative and significant. Observe that only Kortrijk converges with respect to neighboring regions on the opposite side of the language border and only Leuven is affected by short-run spillovers from districts across the language border. Furthermore observe that while housing prices in more districts (Turnhout, Liège, Ghent and Charleroi) converge with respect to housing prices in the dominant region, the error correction component with respect to neighboring districts is very often small in magnitude and the new districts are also affected by short-run spatial spillovers, which makes it hard to interpret the separate effects in isolation.

Table 6: *Estimation results of district specific house price diffusion equation with Antwerp as the dominant region and using the contiguity criterion to construct the appropriate spatial weights matrices (1973Q1-2011Q3)*

District	$\hat{\phi}_{i,0}$	$\hat{\phi}_{i,s_o}$	$\hat{\phi}_{i,s_c}$	$\Delta p_{i,t-1}$	$\Delta \bar{p}^{so}_{i,t-1}$	$\Delta \bar{p}^{sc}_{i,t-1}$	$\Delta p_{0,t}$	Lag order: $\Delta p_{i,t-1}$	$\Delta \bar{p}^{so}_{i,t-1}$	$\Delta \bar{p}^{so}_{i,t-1}$	WH	R^2_i
Antwerp				-0.27***	0.290***			1	1			0.291
				(0.085)	(0.09)							
Mechelen	-0.13**			-0.44***	0.496***		0.281**	3	4		-0.834	0.453
	(0.055)			(0.102)	(0.134)		(0.114)					
Turnhout	-0.10***			-0.39***	0.541***		0.311***	2	2		0.42	0.387
	(0.037)			(0.087)	(0.124)		(0.096)					
Brussels	-0.15***			-0.26***	0.145	0.112	0.336***	4	1	1	0.482	0.544
	(0.043)			(0.103)	(0.091)	(0.093)	(0.066)					
Leuven	-0.25***			-0.36***	-0.1	0.296**	0.340***	2	4	1	0.361	0.506
	(0.062)			(0.078)	(0.112)	(0.141)	(0.092)					
Nivelles	-0.19***			-0.19**	0.243*	0.093	0.295***	1	1	1	-1.587	0.336
	(0.05)			(0.1)	(0.146)	(0.118)	(0.09)					
Bruges	-0.14**	-0.20**		-0.22***	0.047		0.427***	1	1		-0.254	0.394
	(0.057)	(0.084)		(0.07)	(0.102)		(0.101)					
Kortrijk	-0.14***		-0.21***	-0.06	-0.01	-0.14	0.186**	1	1	1	0.318	0.286
	(0.03)		(0.076)	(0.097)	(0.095)	(0.089)	(0.084)					
Veurne		-0.59***		-0.11	-0.13		0.017	1	1		0.589	0.35
		(0.107)		(0.092)	(0.184)		(0.141)					
Dendermonde		-0.42***		-0.06	-0.15		0.166**	1	1		0.247	0.317
		(0.079)		(0.09)	(0.121)		(0.071)					
Ghent	-0.08**			-0.40***	0.255*		0.211	1	1		1.614	0.276
	(0.034)			(0.082)	(0.13)		(0.131)					
Oudenaarde		-0.24***		-0.29***	0.152	0.004	0.135	1	1	1	-0.865	0.289
		(0.064)		(0.094)	(0.196)	(0.136)	(0.143)					
Charleroi	-0.07***			-0.30***	0.207**		0.223***	2	2		-0.345	0.442
	(0.021)			(0.09)	(0.105)		(0.067)					
Tournai		-0.32***		-0.33***	0.246**	0.104	0.259***	1	1	1	0.176	0.486
		(0.081)		(0.085)	(0.117)	(0.106)	(0.075)					
Liège	-0.11***			-0.26***	0.246***	0.053	0.045	1	1	1	-1.107	0.351
	(0.028)			(0.09)	(0.083)	(0.08)	(0.077)					
Verviers	-0.12***	-0.40***		-0.17*	0.087		0.395***	1	1		-1.314	0.462
	(0.044)	(0.095)		(0.093)	(0.118)		(0.114)					
Hasselt		-0.30***		-0.30***	-0.07	0.054	0.233**	1	1	1	0.127	0.412
		(0.072)		(0.102)	(0.099)	(0.093)	(0.105)					
Arlon	-0.14***	-0.28***		-0.25***	0.262		0.394***	2	1		-0.044	0.405
	(0.052)	(0.093)		(0.091)	(0.167)		(0.129)					
Dinant	-0.11**	-0.27***		-0.17**	0.168		0.001	1	2		-0.731	0.369
	(0.048)	(0.075)		(0.074)	(0.187)		(0.129)					
Namur	-0.14***			-0.55***	0.483***		0.175**	2	2		-0.062	0.502
	(0.035)			(0.097)	(0.142)		(0.081)					

Note: Standard errors are shown in parentheses. ***, ** and * signifies that the test rejects the null at respectively the 1, 5 and/or 10% level. All regressions include an intercept term, 3 seasonal dummies and an additional dummy variable to account for the change in the classification process that was implemented by Statistics Belgium at 2005Q1.

While we have already shown that the estimates are not sensitive with respect to the sample period under consideration, it is also important to investigate the robustness of our results with respect to the spatial weights matrices. As was already mentioned the results are conditional upon the specification of the spatial weights matrices which are *a priori* defined by the researcher. Therefore, we present the results of our baseline model where the inverse distance criterion was used to construct the spatial weights matrices in table 7. The results presented are very similar to those presented in tables 5 and 6 and thus require little additional explanation.

5.5 Generalized spatio-temporal Impulse Response Functions

Although the coefficient estimates presented in tables 5, 6 and 7 shed light on the statistical relationships among housing prices in the different Belgian districts, the estimated coefficients are ingredients of a large and complex system of interactions and feedback mechanisms that can hardly be interpreted in isolation. Therefore, similar to Holly *et al.* (2011), we use Generalized spatio-temporal Impulse Response Functions (GIRF) to analyze how (idiosyncratic) shocks are propagated over time and space. Figure 2 displays the spatial and temporal diffusion of an idiosyncratic shock (one standard deviation) to the dominant region using the estimates from the baseline model presented in table 5.[10]

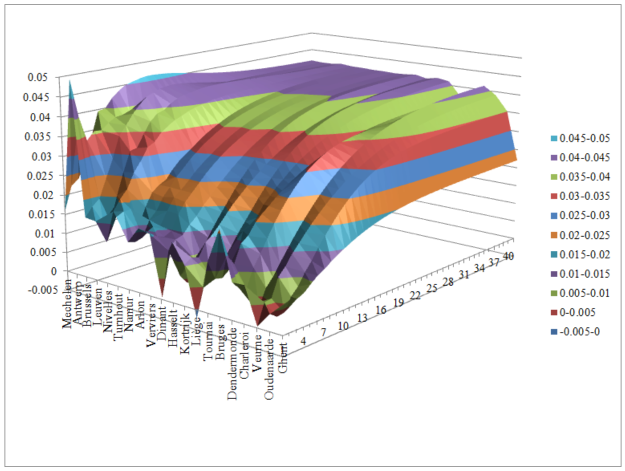
[10] In appendix B we present the GIRFs for each district separately together with their bootstrapped (5000 replications) confidence intervals.

Table 7: *Estimation results of district specific house price diffusion equation with Antwerp as the dominant region and using the inverse distance criterion to construct the appropriate spatial weights matrices (1980Q1-2011Q3)*

									Lag order:				
District	$\hat{\phi}_{i,0}$	$\hat{\phi}_{i,s_o}$	$\hat{\phi}_{i,s_c}$	$\Delta p_{i,t-1}$	$\Delta \bar{p}^{so}_{i,t-1}$	$\Delta \bar{p}^{sc}_{i,t-1}$	$\Delta p_{0,t}$	ΔGDP_t	$\Delta p_{i,t-1}$	$\Delta \bar{p}^{so}_{i,t-1}$	$\Delta \bar{p}^{so}_{i,t-1}$	WH	R^2_i
Antwerp				-0.29***	0.349***			0.907***	1	1			0.329
				(0.093)	(0.108)			(0.347)					
Mechelen	-0.21***			-0.26**	0.406***		0.398***	-0.25	1	1		-0.291	0.445
	(0.066)			(0.113)	(0.144)		(0.114)	(0.398)					
Turnhout		-0.19***		-0.26***	0.337***		0.363***	0.556	2	1		0.87	0.469
		(0.05)		(0.087)	(0.126)		(0.097)	(0.417)					
Brussels	-0.15***			-0.08	-0.01	0.140**	0.312***	0.445*	4	1	1	0.783	0.598
	(0.046)			(0.088)	(0.086)	(0.07)	(0.074)	(0.238)					
Leuven	-0.25***			-0.27***	-0.16	0.094	0.335***	0.06	1	4	1	0.354	0.479
	(0.076)			(0.089)	(0.128)	(0.176)	(0.116)	(0.384)					
Nivelles	-0.17***			-0.33***	0.191	0.300***	0.215***	1.05***	1	1	1	-2.212**	0.515
	(0.048)			(0.076)	(0.126)	(0.094)	(0.077)	(0.34)					
Bruges	-0.16***	-0.21**		-0.20**	-0.03		0.419***	1.06***	1	1		1.107	0.523
	(0.047)	(0.082)		(0.088)	(0.101)		(0.094)	(0.385)					
Kortrijk	-0.14***		-0.17*	-0.12	-0.13	-0.04	0.212**	0.554	1	1	1	1.820*	0.352
	(0.032)		(0.097)	(0.113)	(0.104)	(0.107)	(0.098)	(0.381)					
Veurne		-0.71***		-0.03	-0.38**		0.001	0.442	1	1		1.045	0.391
		(0.134)		(0.095)	(0.177)		(0.141)	(0.498)					
Dendermonde		-0.32***		-0.05	-0.15		0.182**	0.616	1	1		1.214	0.375
		(0.077)		(0.106)	(0.122)		(0.081)	(0.435)					
Ghent				-0.59***	0.524***		0.133	0.614	2	2		0.96	0.337
				(0.109)	(0.133)		(0.161)	(0.441)					
Oudenaarde		-0.45***		-0.26***	-0.1	-0.05	0.075	-0.24	1	1	1	0.039	0.419
		(0.091)		(0.092)	(0.22)	(0.138)	(0.144)	(0.536)					
Charleroi		-0.28***		-0.17*	-0.18		0.195***	0.29	1	1		1.045	0.536
		(0.045)		(0.093)	(0.114)		(0.069)	(0.271)					
Tournai		-0.22**		-0.38***	0.296*	0.011	0.249***	0.162	2	4	1	-1.495	0.598
		(0.091)		(0.117)	(0.169)	(0.09)	(0.084)	(0.369)					
Liège		-0.16**	-0.22***	-0.18*	0.008	-0.04	-0.04	0.547	1	1	1	-0.609	0.556
		(0.078)	(0.056)	(0.104)	(0.102)	(0.076)	(0.082)	(0.337)					
Verviers	-0.17***			-0.38***	0.302**		0.290**	0.459	2	1		-0.204	0.475
	(0.044)			(0.09)	(0.125)		(0.132)	(0.566)					
Hasselt		-0.31***		-0.33***	-0.1	-0.04	0.298***	0.263	1	1	1	0.017	0.442
		(0.085)		(0.117)	(0.106)	(0.097)	(0.106)	(0.601)					
Arlon	-0.20***	-0.31**		-0.15	0.128		0.324***	0.926	2	1		-0.131	0.428
	(0.058)	(0.133)		(0.127)	(0.178)		(0.125)	(0.705)					
Dinant	-0.09*	-0.47***		-0.18*	0.225		0.003	0.983**	1	1		-0.267	0.439
	(0.049)	(0.157)		(0.109)	(0.202)		(0.124)	(0.457)					
Namur	-0.15***			-0.42***	0.230*		0.254***	0.692*	1	1		0.38	0.476
	(0.034)			(0.089)	(0.122)		(0.093)	(0.379)					

Note: Standard errors are shown in parentheses. ***, ** and * signifies that the test rejects the null at respectively the 1, 5 and/or 10% level. All regressions include an intercept term, 3 seasonal dummies and an additional dummy variable to account for the change in the classification process that was implemented by Statistics Belgium at 2005Q1.

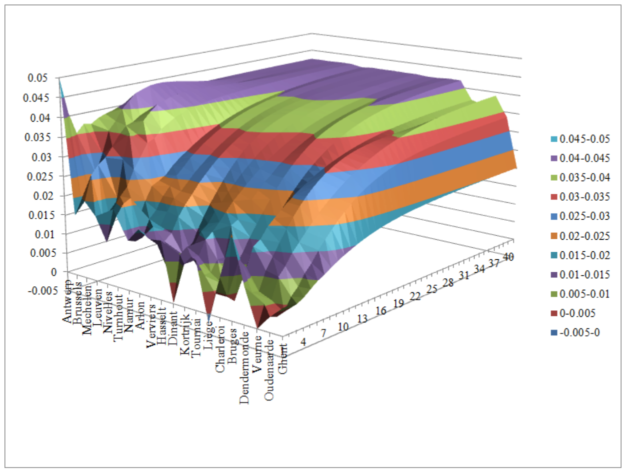
Figure 2: *Generalized spatio-temporal Impulse Response Functions (horizon = 40) for our baseline estimates (1980Q1-2011Q3)*



Note: the different districts are ordered according to their respective averages over time.

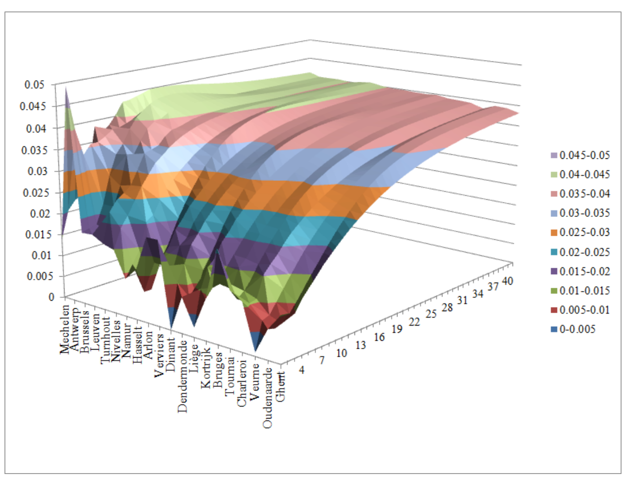
The simulated GIRFs clearly provide evidence for the existence of *the ripple effect hypothesis* in Belgium. Shocks to the dominant region are only slowly propagated to other regions. A quick glance suggests that it takes time (approximately 10 years) before a single shock is fully absorbed by the remaining regions, which implies that there are frictions that prevent housing markets to converge to their new equilibrium immediately. Also observe that districts located along the north-south axis of Belgium, which constitutes the economic spine of the country, converge much faster with respect to Antwerp than districts in the more peripherally located eastern and western parts of the country. Given that the initial convergence process follows the line Antwerp-Namur, and thus crosses the language border, our results suggest that the regions along this axis are sufficiently integrated to overcome their linguistic differences. Our intuition for this result is that the bilingual Brussels Capital Region, a main driver of the Belgian economy, is likely to be the cement that connects the district Brussels to the French-speaking district of Nivelles (and subsequently to Namur) and thus spans across the linguistic border. Once the convergence process has initially occurred along the north-south axis districts that are located in the eastern and western parts of Belgium converge with respect to their neighbors. Although this cannot explicitly be seen from the graph presented in figure 2, the estimated error correction coefficients presented in table 5 suggested that this convergence process occurs within each linguistic region. The GIRFs for our robustness checks, which display very similar patterns, are presented in figures 3 and 4.

Figure 3: *GIRFs using the full-sample (1973Q1-2011Q3) estimates and the contiguity criterion to construct the appropriate spatial weights matrices*



Note: the different districts are ordered according to their respective averages over time.

Figure 4: *GIRFs using the reduced sample (1980Q1-2011Q3) estimates and the inverse distance criterion to construct the appropriate spatial weights matrices*



Note: the different districts are ordered according to their respective averages over time.

5.6 The role of language

The estimation results presented in tables 5, 6 and 7 and their corresponding GIRFs presented in figures 2, 3 and 4 suggest that peripherally located districts in the eastern and western parts of Belgium are far more integrated with neighboring districts within their linguistic area than with neighboring districts located across the linguistic border. Although we cannot explicitly test using our econometric model that language as such is the main driver, one might be concerned that there are other factors, which are correlated with language, that might drive our results.

One might, for example, be concerned that the different peripheral districts in our dataset have different economic structures which prohibit the integration of housing prices with their neighbors across the linguistic border. In table C.1 in appendix C we present data that was provided by the National Bank of Belgium (NBB) concerning the number of people employed in different sectors of the economy in 2011 at the provincial level. The figures indicate that there are only minor differences for the different provinces, which suggest that there is no reason to believe that differences in economic structure explain our results.

One might also be concerned that our results are caused by differences in the composition of sales. To this end we performed an additional robustness check by estimating the same model for dwellings only. The estimation results are presented in table C.2 and the GIRFs are plotted in figure C.1, which are both in appendix C, again are very to those presented previously. We observe a strong interdependence among districts located along the north-south axis of Belgium, while peripheral districts in the east and west almost exclusively converge with respect to their neighbors on the same side of the language border. This suggests that the results presented earlier are not driven by changes in composition. Although the results presented in this section do not prove that linguistic differences cause the low degree of interdependence among housing prices in peripheral districts, it is at least remarkable that housing prices in a small country such as Belgium (approximately 30,000km2, compared to 245,000km^2 for the UK and even 9,000,000km^2 for the US) partially follow these linguistic patterns.

6 Concluding remarks

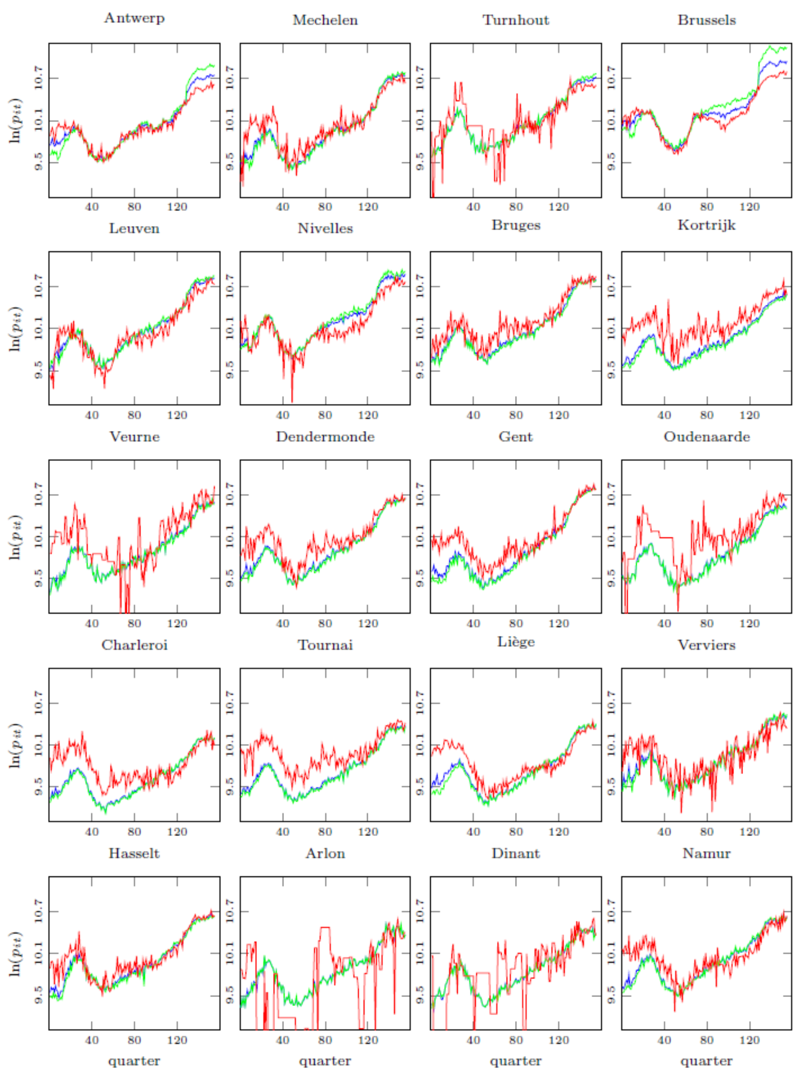
In this paper we have assessed the validity of *the ripple effect hypothesis* for Belgium, and have especially focused on the role of linguistic border that divides Belgium into two large linguistic regions. Simple descriptive statistics presented in table 2 revealed that the degree of interdependence among regional house prices is higher between districts within a linguistic region, than between districts located in a different language region. After that we determined a suitable candidate, Antwerp, for the dominant region in our model in section 5.2 we estimated a flexible econometric model where we allow for a full set of possible interactions and error correction mechanisms between different regions. The spatial weights matrices used to construct spatially weighted averaged additionally take into account the unique federal structure of Belgium. The results presented in table 4 and figure indicate that regional housing prices exhibit a distinct spatial pattern where housing prices in districts located along the centrally located north-south axis of Belgium are more highly integrated among each other and converge faster with respect to the dominant region, Antwerp. In a second wave of this convergence process, housing prices in the more peripherally located eastern and western parts of Belgium converge almost exclusively with respect to neighboring districts in the same linguistic region, which suggests that there are strong effects from the language border in these areas. Although the current econometric framework does not allow to explicitly test whether the observed effects are solely driven by linguistic differences, the results presented in section 5.6 suggest that the economic structure on both sides of the language border is not that different, which makes our case stronger.

A first takeaway from this paper is that, similar to the results found in many studies in the UK, regional housing prices in Belgium exhibit a distinct spatial pattern over time that is consistent with the ripple effect hypothesis. The second, perhaps more important, takeaway from this paper is that linguistic differences between different regions might result into observed border effects. This suggests that merely removing institutional differences is unlikely to be sufficient to achieve full integration between different regions. Although we do not claim that the results from the current study can easily be extrapolated to other regions/countries and areas of study, the results may provide valuable insights for, for example, the European integration process. While the EU has been aiming at (and continues to aim at) removing institutional differences between its different members states over the past decades, differences in economic variables continue to persist between the different countries. It is obvious that the different member states of the EU continue to be heterogeneous with respect to factors such as language and culture.

Appendices

A Raw data (1973Q1-2011Q3) and the resulting ADF-tests

Figure A.1: *Overview of the raw data (1973Q1-2011Q3)*



Note: the red, green and blue series represent respectively apartments, dwellings and the mix-adjusted series constructed using the appropriate weighting procedure described in section 3.

Table A.1: *Augmented Dickey-Fuller (ADF) test statistics (1973Q1-2011Q3)*

	Mix-adjusted		Dwellings		Apartments	
District	ln(.)	Δln(.)	ln(.)	Δln(.)	ln(.)	Δln(.)
Antwerp	-0.95	-9.50***	-0.85	-7.72***	-1.34	-13.0***
Arlon	-1.37	-11.6***	-1.32	-11.4***	-4.09***	-3.17**
Bruges	-0.89	-10.2***	-0.82	-9.20***	-1.71	-17.3***
Brussels	-1.72	-2.78*	-1.25	-7.09***	-1.4	-2.84*
Charleroi	-1.5	-3.83***	-1.46	-3.93***	-1.24	-14.0***
Dendermonde	-0.96	-11.1***	-0.91	-10.8***	-1.66	-13.3***
Dinant	-1.33	-11.0***	-1.25	-11.2***	-3.58**	-3.30**
Ghent	-1.1	-12.0***	-1.11	-12.0***	-1	-11.4***
Hasselt	-0.97	-10.0***	-1.07	-10.3***	-1.67	-8.55***
Kortrijk	-0.84	-10.9***	-0.68	-11.0***	-1.85	-11.9***
Leuven	-1.1	-11.6***	-1.12	-10.6***	-1.32	-5.58***
Liège	-1.17	-3.26**	-1.15	-2.75*	-0.713	-11.5***
Mechelen	-1.37	-4.80***	-1.07	-9.74***	-1.48	-10.5***
Namur	-0.92	-3.30**	-1.07	-3.16**	-1.25	-14.6***
Nivelles	-1.08	-8.61***	-1.04	-4.08***	-2.33	-13.0***
Oudenaarde	-1.03	-10.6***	-1.04	-11.5***	-3.12*	-4.98***
Tournai	-1.23	-2.84*	-1.21	-2.91**	-1.73	-14.2***
Turnhout	-1.15	-10.9***	-1.05	-4.59***	-2.45	-6.63***
Verviers	-0.99	-12.6***	-1.03	-12.5***	-2.01	-11.5***
Veurne	-1.37	-12.0***	-1.45	-11.7***	-1.63	-4.86***

Note: the lag orders are selected using the SIC-criterion. A trend is added for the series in levels. ***, ** and * signify the test rejects the null at respectively the 1, 5 and/or 10% level.

B Choice of the dominant region

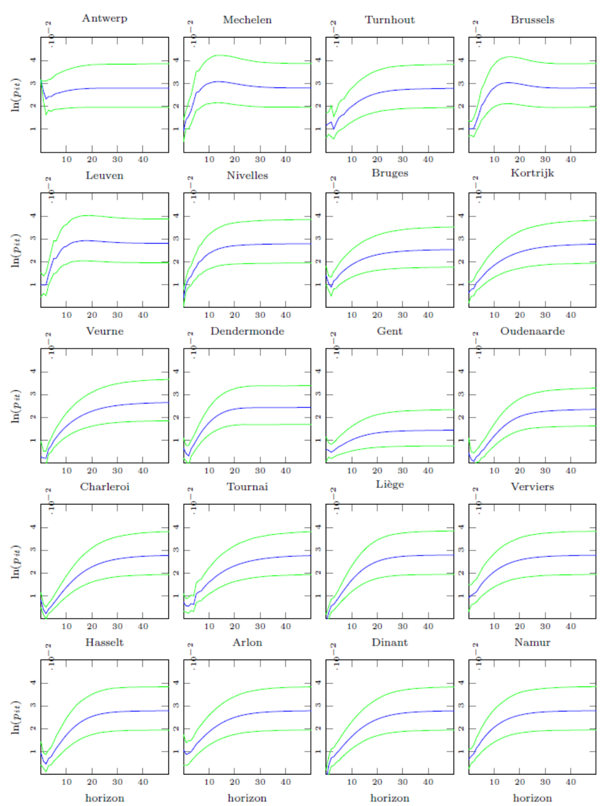
Table B.1: *Results Granger & Lin (1995) long-run causality tests*

	ASSUMED DOMINANT REGION (region i)																			
Region j	ANT	MEC	TUR	BRX	LEU	NIV	BRU	KOR	VEU	DEN	GHE	OUD	CHA	TOU	LIE	VER	HAS	ARL	DIN	NAM
ANT	n/a				*													*		
MEC	***	n/a		**	***						**									
TUR	***		n/a	***	***	**	***	***	**	**	**	**	**	***	**	***	*	**	**	***
BRX	**			n/a																
LEU	***			***	n/a													*	*	
NIV	***			*	***	n/a	**													*
BRU	**	*		**	***		n/a													
KOR	**	*		**	***	**	*	n/a												**
VEU	***	***	***	***	***	***	***	**	n/a	**	**	*		**	*	**	**	**	**	***
DEN	***	***		**	***	**	**			n/a	**									
GHE											n/a				*					
OUD	***	**		***	***	**	**	*		**	**	n/a		**	**	**		**		***
CHA	*			*	**			**					n/a	**		**		**		**
TOU	**			**	***									n/a		*				**
LIE	**	**	*	**	***	**	**	*		**				*	n/a	*		**	*	***
VER	***	**		***	***	**	*	*		*				*		n/a				***
HAS	***	*	**	***	***	**	***	**	**	***	*	*	*	**	**	**	n/a	**	**	***
ARL	***		*	***	**	***	*	***		**			***	***	***	***		n/a	***	***
DIN	***		***	***	**	***	**	***		*			**	***	**	***		***	n/a	***
NAM	*			*	***															n/a

Note: the *, ** and *** indicate that the assumed dominant region i is long-run forcing upon region j at the 10, 5 and/or 1% level, respectively.

C Generalized spatio-temporal Impulse Response Functions and their bootstrapped confidence intervals (horizon = 50) for the baseline model

Figure C.1: *Generalized spatio-temporal Impulse Response Functions and their Bootstrapped Confidence Intervals (horizon = 50) for the baseline model presented in table 5*



Note: the depicted GIRFs and their confidence intervals are based upon the baseline estimates presented in table 5 and using the procedure described by Holly *et al.* (2011) to perform the bootstrap procedure. The other series are available from the authors upon request.

Table D.1: *Employment shares for the different provinces and sectors for the year 2011*

Sector	BCR	Antwerp	Limburg	EF	FB	WF	WB	Hainaut	Liège	Lux.	Namur	Belgium
Construction	3%	6.14%	7.54%	7.82%	4%	7.37%	5.39%	6.8%	7.55%	9.16%	6.74%	6.15%
Culture, recreation and other services	5.29%	3.89%	4.07%	3.88%	3.89%	4.08%	4.66%	4.24%	4.73%	5.16%	5.46%	4.35%
Minerals, manufacturing and utilities	4.42%	15.67%	17.09%	15.76%	9.5%	18.57%	13.61%	13.71%	13.07%	10.61%	8.77%	12.97%
Financial services	8.73%	2.33%	1.42%	1.83%	1.97%	1.46%	1.93%	1.52%	1.73%	1.19%	1.88%	2.86%
Trade and exploitation real estate	0.82%	0.46%	0.34%	0.45%	0.34%	0.57%	0.7%	0.59%	0.56%	0.29%	0.44%	0.53%
Trade, logistics and catering industry	19.57%	23.65%	20.76%	20.19%	27.57%	21.31%	21.37%	22.04%	21.66%	20.92%	20.06%	21.9%
ICT	4.57%	2.04%	1.3%	1.36%	4.35%	1.16%	3.22%	1.13%	1.42%	0.89%	1.57%	2.27%
Agriculture, forestry and fisheries	0.01%	1.06%	2.06%	1.61%	1.05%	2.63%	0.86%	1.32%	1.12%	4.05%	2.03%	1.34%
Pub. man., edu. and health serv.	34.57%	24.96%	27.78%	29.08%	24.93%	26.94%	26.88%	35.15%	33.78%	36.08%	38.49%	29.9%
Services to business	18.99%	19.76%	17.58%	17.97%	22.36%	15.86%	21.33%	13.46%	14.34%	11.61%	14.52%	17.7%

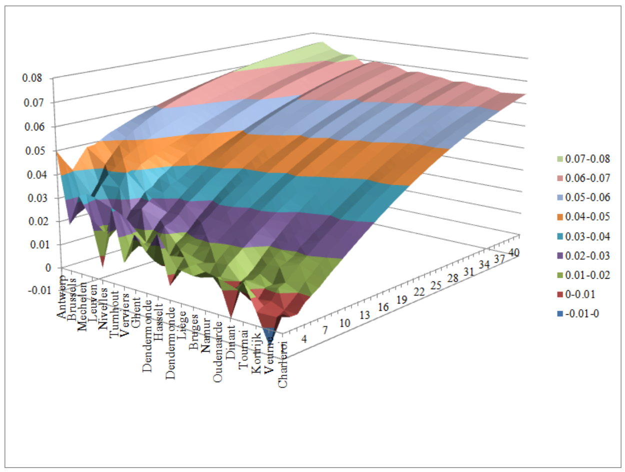
Table D.2: *Estimation results of district specific house price diffusion equation (only dwellings) with Antwerp as the dominant region and using the contiguity criterion to construct the appropriate spatial weights matrices (1980Q1-2011Q3)*

District	$\hat{\phi}_{i,0}$	$\hat{\phi}_{i,s_o}$	$\hat{\phi}_{i,s_c}$	$\Delta p_{i,t-1}$	$\Delta\bar{p}^{so}_{i,t-1}$	$\Delta\bar{p}^{sc}_{i,t-1}$	$\Delta p_{0,t}$	$\Delta \mathrm{GDP}_t$	Lag order: $\Delta p_{i,t-1}$	$\Delta\bar{p}^{so}_{i,t-1}$	$\Delta\bar{p}^{so}_{i,t-1}$	WH	R^2_i
Antwerp				-0.29***	0.354***			0.695**	1	3			0.589
				(0.101)	(0.095)			(0.299)					
Mechelen	-0.32***	-0.17**		-0.18*	0.171		0.662***	-0.32	1	1		-1.224	0.594
	(0.082)	(0.07)		(0.107)	(0.117)		(0.115)	(0.318)					
Turnhout		-0.17***		-0.31***	0.294***		0.632***	0.685*	2	1		0.824	0.563
		(0.051)		(0.093)	(0.108)		(0.098)	(0.411)					
Brussels	-0.12***			-0.12**	0.144*	0.083	0.398***	0.409	1	4	1	-0.312	0.763
	(0.043)			(0.052)	(0.085)	(0.065)	(0.086)	(0.291)					
Leuven	-0.17***		-0.25***	-0.18*	-0.03	0.081	0.506***	-0.1	1	4	1	0.208	0.601
	(0.052)		(0.098)	(0.111)	(0.116)	(0.139)	(0.116)	(0.393)					
Nivelles	-0.10***			-0.48***	0.386***	0.235***	0.106	1.60***	1	1	1	-2.457**	0.519
	(0.032)			(0.087)	(0.149)	(0.066)	(0.12)	(0.483)					
Bruges	-0.11***	-0.21*		-0.28***	0.009		0.250*	1.20***	1	1		-2.460**	0.516
	(0.029)	(0.115)		(0.104)	(0.123)		(0.132)	(0.413)					
Kortrijk	-0.07***			-0.34***	0.08	-0.03	0.106	0.504	2	2	1	0.229	0.384
	(0.022)			(0.116)	(0.124)	(0.086)	(0.116)	(0.355)					
Veurne		-0.83***		0	-0.36**		-0.13	0.427	1	1		0.179	0.433
		(0.143)		(0.1)	(0.184)		(0.169)	(0.494)					
Dendermonde		-0.23***		-0.12	-0.11		0.339***	0.785*	2	1		-0.229	0.473
		(0.055)		(0.107)	(0.098)		(0.096)	(0.412)					
Ghent	-0.10***			-0.38***	0.327**		0.458**	0.209	2	1		-1.187	0.396
	(0.037)			(0.09)	(0.13)		(0.2)	(0.426)					
Oudenaarde		-0.32***	-0.17**	-0.26***	-0.14	-0.06	0.289*	-0.28	1	1	1	-0.576	0.467
		(0.072)	(0.084)	(0.093)	(0.186)	(0.142)	(0.155)	(0.553)					
Charleroi		-0.34***		-0.21**	-0.04		0.149	0.404	1	1		0.806	0.522
		(0.062)		(0.097)	(0.124)		(0.098)	(0.309)					
Tournai		-0.21**		-0.39***	0.347**	-0.07	0.312***	0.051	2	4	1	-0.724	0.644
		(0.087)		(0.109)	(0.143)	(0.126)	(0.111)	(0.389)					
Liège	-0.07***	-0.18***		-0.40***	0.113	0.190**	0.266***	0.473	2	1	1	-1.607	0.619
	(0.019)	(0.058)		(0.09)	(0.098)	(0.076)	(0.096)	(0.377)					
Verviers	-0.14***			-0.37***	0.229*		0.197	0.688	2	1		0.836	0.419
	(0.035)			(0.098)	(0.117)		(0.16)	(0.625)					
Hasselt		-0.25***		-0.25*	0.023	-0.14	0.514***	-0.17	1	1	1	0.858	0.428
		(0.087)		(0.132)	(0.135)	(0.098)	(0.129)	(0.67)					
Arlon	-0.12***	-0.30**		-0.19	0.161		0.117	1.51**	2	1		-0.283	0.404
	(0.045)	(0.129)		(0.12)	(0.189)		(0.166)	(0.657)					
Dinant		-0.54***		-0.21**	0.415**		0	1.06**	1	1		-0.297	0.446
		(0.147)		(0.107)	(0.179)		(0.157)	(0.5)					
Namur	-0.06***			-0.34***	0.223*		0.310**	0.905**	1	1		0.338	0.462
	(0.025)			(0.085)	(0.132)		(0.123)	(0.392)					

Note: Standard errors are shown in parentheses. ***, ** and * signifies that the test rejects the null at respectively the 1, 5 and/or 10% level. All regressions include an intercept term, 3 seasonal dummies and an additional dummy variable to account for the change in the classification process that was implemented by Statistics Belgium at 2005Q1.

D Additional tables and figures for section concerning the role of language

Figure D.1: *GIRFs using the reduced-sample (1980Q1-2011Q3) estimates for dwellings only and the contiguity criterion to construct the appropriate spatial weights matrices*



Note: the different districts are ordered according to their respective averages over time.

Chapter II

Spatial Arbitrage in Belgian Border Regions

1 Introduction

It has been widely recognized that the housing commodity is characterized by a number of peculiar features, such as locational fixity, durability and heterogeneity.[1] Despite these features and the substantial costs associated with moving, it is nonetheless to be expected that differences in house prices are limited between locations that are close to each other when barriers to mobility are limited as a result of (spatial) arbitrage (Glaeser & Gyourko, 2007). In the case of national borders, however, differences in mortgage markets and the language spoken on both sides of the border potentially limit the scope for- and desirability of arbitrage. In a seminal paper that was published in the *American Economic Review*, McCallum (1995) shows that "*even the relatively innocuous Canada-US border matters substantially for international trade patterns.*" In another seminal paper Engels & Rogers (1996, *AER*) find that "*crossing the border is equivalent to 1,780 miles of distance between cities in terms of the price dispersion of similar goods.*" While the aforementioned studies focus explicitly on international trade patterns, border effects have also been reported in housing markets. In a recent pa-

[1] I would like to thank Erik Buyst, Frank Verboven, Maarten Goos, Jan Rouwendal, Jan Mutl, Sven Damen, Geert Goeyvaerts, Frank Vastmans and the participants of the European Network for Housing Research conference (July, 2015) and the (other) members of the Flemish Policy Research Centre Housing (Dutch: *Steunpunt Wonen*) for valuable comments and suggestions on earlier drafts of this paper. I would furthermore like to thank ERA Belgium, *Statistics Belgium*, *Statistics Netherlands*, and the Dutch association of real estate agents (NVM, Dutch: *Nederlandse Vereniging van Makelaars and Taxateurs in onroerende goederen*) for providing the data that was necessary to carry out the analyses performed in this paper.

per, Micheli *et al.* (2014) show, using the results from different estimation strategies, that the ask prices of comparable housing drop by about 16% when crossing the Dutch-German border. The authors argue that "*Dutch households might be paying as much as 26% higher house prices to live among their own people.*" These and other studies thus clearly show that borders between countries limit the scope for arbitrage.

In this paper we investigate whether houses that are more proximate to the Belgian-Dutch border are more expensive as a result of spatial arbitrage. Both Belgium and the Netherlands are among the founding fathers of the European Union and the Benelux and share a long common history.[2] The northern Flemish part of Belgium, which borders the Netherlands for the most part and will thus be of special interest in this paper even shares a common language with the Netherlands.[3] Despite that both countries have largely been subjected to similar monetary and macro-economic policies and shocks, the evolution of housing prices in both countries has differed substantially in recent decades. While housing prices in the Netherlands have boomed from the mid-1990s onwards as a result of, among other things, an increase in the share of interest-only mortgages (Rouwendal, 2007), Belgian housing prices only started to increase strongly from the mid-2000s due to a more generous fiscal treatment of owner-occupiers (Damen *et al.*, 2016). This strong initial increase in housing prices in the Netherlands, combined with the fact that housing prices in the Netherlands already exceeded those in Belgium, led to a large discontinuity in housing prices at the border which has led to opportunities for spatial arbitrage. In this paper we investigate whether the prices of houses in Belgium that are located in closer proximity to the Belgian-Dutch border are higher as a result of spatial arbitrage. The intuition is that Dutch households, who moved (just) across the border to benefit from the lower housing prices in Flanders, increased the demand and prices of Belgian properties close to the border. We therefore use a large sample of detailed individual transaction data that was provided by a large franchise system of real estate agencies in Belgium and employ the well-known (spatial) hedonic pricing model (Rosen, 1974). Our results suggest

[2]Belgium seceded from the Netherlands in 1830.

[3]Belgium is a federal state that is divided into three regions and three communities, that exist next to each other. Its two largest regions are the Dutch-speaking region of Flanders in the north and the French-speaking southern region of Wallonia. The Brussels-Capital Region, officially bilingual, is a mostly French-speaking enclave within the Flemish Region.

that at the start of the sampling period (2003), the price difference between a property located at the border and a property located 15 kilometers from the border was about 13 percent. The boom in Belgian housing markets since 2005, combined with the subsequent collapse of the Dutch housing market after the global financial crisis of 2007-2008, furthermore suggests that the magnitude of the spatial arbitrage effect might not have been constant over time. In the empirical application of this paper, we therefore allow for interaction effects between the distance to the border and the year of sale. Our results indicate that the magnitude of the spatial arbitrage effect although dwellings located 15 kilometers from the border were 13% less expensive in 2003, this effect has decreased to only 2% in 2015. Since the Belgian-Dutch border stretches out over 460 kilometers, we also investigate whether the estimated effects are constant across space. Our results indicate that the estimated spatial arbitrage effect is especially strong in the eastern parts of Flanders, where sparsely populated Flemish regions border Dutch cities. In a final extension, we also examine whether Dutch buyers pay a premium compared to their Belgian counterparts. We indeed find evidence that Dutch buyers pay a premium of about 12%. This premium, however, has also been declining over time which seems to suggest that these premiums are likely to be the result of anchoring bias.

The remainder of this paper is organized as follows. In section 2 we provide an overview of the literature that is related to the current study. In the third section we examine the developments in housing prices and markets in Belgium and the Netherlands in recent decades. In section 4 we present the data that is used in the subsequent empirical analyses. In section 5 we lay out our empirical framework and in section 6 we present the baseline results. In section 7, we report the results of a battery of robustness checks and extensions. Finally, section 8 concludes.

2 Literature review

Border effects have mainly been studied in the context of international trade patterns. In a seminal article published in the *American Economic Review* (1995) John McCallum shows that the border between Canada and the United States, two countries that are similar in terms of culture, language and institutions, continues to have a decisive impact on continental trade patterns. In another seminal paper that was published in the *AER*, Engel

& Rogers (1996) find that crossing the border is equivalent to 1,780 miles of distance between cities in terms of price dispersion of similar goods. Following these seminal contributions a large and well-encompassing strand of literature has developed where researchers look at the existence, magnitude and explanations of border effects. Well-known examples include Parsley & Wei (2001), Anderson & van Wincoop (2003) and Gorodnichenko & Tesar (2009).

Although the literature on border effects originated in the context of international trade patterns, border effects have also been studied in the context of housing markets and housing prices. Cheshire & Magrini (2009), for example, find that national borders still have a significant impact upon property prices. They furthermore show that this is due to the fact that cities within the Union still form national urban systems rather than a single European-wide system. These results are also confirmed by Jacobs-Crisioni & Koomen (2015), who show that national borders still affect the spatial urban pattern in northern Europe. While the aforementioned studies look at the effects of national borders, there are also a number of studies that examine the effects of linguistic/cultural borders *within* a country. Goffette-Nagot *et al.* (2009), for example, examine the spatial variation of land prices in Belgium and find that the linguistic border acts as a strong barrier in the spatial pattern of land prices. De Bruyne & Van Hove (2013) also observe that there are large differences in housing prices between the northern (Flanders) and the southern (Wallonia) part of Belgium in their analysis of the determinants of housing prices. They also examine whether the determinants of housing prices differ between the regions by splitting up their sample and re-estimating the model. Their results suggest that the effect of various explanatory variables on housing prices is different in the two regions. In a recent article Micheli *et al.* (2014) show, by combining German and Dutch real estate datasets, that the listing prices of observationally equivalent properties drop by about 16% when crossing the Dutch-German border. Given that the price discounts in Germany are substantially larger, the authors interpret their findings as indicating that "*the willingness of Dutch households to pay up to 26% higher housing prices to live among the Dutch.*" While the aforementioned studies investigate the "jump at the border", the current paper examines whether housing prices "closer to the border" are higher as a result of spatial arbitrage.[4] If Dutch households indeed prefer to live among the Dutch or close to the Netherlands, we would expect that housing prices closer to the border, *ceteris paribus*, are higher.

The current paper also relates to a stand of literature that investigates spatial and temporal spillovers in housing markets and housing prices across countries. The majority of studies in this context have focused on the European Union, where the national economies are perhaps even more integrated than those of the US and Canada. Stevenson (2004), for example, examines house price diffusion patterns *within* the Republic of Ireland and *between* the Republic and Northern Ireland. His results support the view that the Northern Irish market is more linked with the housing market in the Republic than with the rest of the UK. Vansteenkiste & Hiebert (2011) estimate a global VAR for three housing demand variables and seven Euro area countries and find limited evidence for house price spillovers in the Euro area. While these studies explicitly focus at national borders, there are also some studies that focus on the role of linguistic borders in housing markets. In a recent study Helgers & Buyst (2016) examine whether the linguistic border has an effect on the (long-run) convergence and the (short-run) spatial and temporal diffusion of housing prices in Belgium. Their results suggest that the border plays an ambiguous role. In the current study we also examine whether the spatial arbitrage effect is constant across time and space.

Our work also relates to a strand of literature that has focused on the role of language on economic outcomes. It has been well-documented in the international trade literature (e.g. Melitz, 2008) that the sharing of a language increases the volume of traded goods and services between two countries. Although the emergence of nation-states has led to sharp discontinuities in the language spoken on both sides of the border for many European countries, Flanders and the Netherlands share a common language. We would therefore expect that the border effect at the Flemish-Dutch border is smaller than the border effect at, for example, the Dutch-German border as documented by Micheli *et al.* (2014).

Despite that language obviously is one of the most obvious cultural discrepancies between countries, there are also other cultural factors that play a role.

[4] Our empirical model can in principle be easily extended to also estimate the jump at the border when individual transaction data on both sides of the border is available.

Falck *et al.* (2012), for example, show in study of cross-regional migration patterns in Germany between 2000 and 2006 that historical dialect similarity has an important effect and increases cross-regional migration flows. Besides the aforementioned cultural differences, there are also differences in spatial planning between countries. Tennekes *et al.* (2015), for example, note that "*it is still relatively easy to see the lines of the border when crossing from the Netherlands to Belgium, not only due to road signage and illumination but also due to differences in dwelling types and the shape of urban development.*" The authors argue that "*differences in national institutional environments have contributed to differences in the urban morphology of residential areas.*"

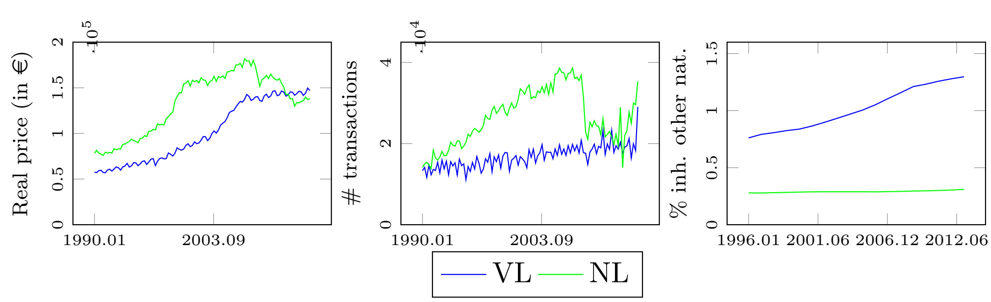
This study finally also relates to a strand of literature that investigates whether certain types of buyers pay a premium over other types of buyers. While countless studies have shown that housing prices depend upon the attributes of houses, these studies have also shown that after controlling for these characteristics, there is not one price but rather a distribution of prices. Besides methodological issues, a potential explanation for this finding is that housing markets are thin markets where prices are negotiated and thus affect by the relative bargaining strength of buyers and sellers. While the hedonic pricing framework has become the workhorse model for housing economists, there is no explicit role for bargaining in the traditional model (Harding et al., 2003), since the implicit prices of characteristics are revealed to agents and markets are assumed to be sufficiently thick. While Harding *et al.* (2003) and Ihlanfeldt & Mayock (2009) find evidence that suggests that demographic characteristics of buyers and sellers influence the sales prices of observationally equivalent properties, it has also been hypothesized that *out-of-state* buyers pay a premium because of higher search costs, anchoring bias, and/or a lack of market knowledge. Turnbull & Sirmans (1993) and Watkins (1998), however, do not find evidence that supports this hypothesis. Lambson *et al.* (2004) and Ihlanfeldt & Mayock (2011) though do find strong evidence. Moreover, the results provided by Lambson *et al.* (2004) suggest that these premiums are driven by higher search costs and anchoring bias. In the current paper we use information on the previous address of buyers to investigate whether Dutch buyers pay a premium compared to their Belgian counterparts. Our findings indicate that Dutch buyers pay a premium, that is likely to be the result of anchoring bias. In the next section, we provide an overview of the developments in Belgian and Dutch housing markets in recent decades.

3 Housing Prices and Housing Markets in Belgium and the Netherlands

3.1 A general overview

Vansteenkiste & Hiebert (2011) recently argued that despite that housing is a non-traded good that cannot easily be substituted across geographic areas, co-movement in international housing prices could nevertheless be expected to arise from three different channels, notably (1) *common developments in housing market fundamentals*, (2) *the parallel introduction of capital and mortgage innovations*, and (3) "*[...]housing-specific factors, notable related to some convergence of housing risk premia associated with returns on housing as an asset*" (p. 299). Given that Belgium and the Netherlands are both members of the EMU, with its common monetary policy, housing market fundamentals such as interest rates are similar. Both economies are furthermore close trading partners which suggests that the risk premia should also converge.[5] In figure 1 we plot the evolution of average real house prices and the number of transactions in Belgium and the Netherlands using data from *Statistics Belgium*, *Statistics Netherlands*, and the *NVM*[6].

Figure 1: *Real house prices, # of transactions, and the % of people with the other nationality for Belgium and the Netherlands (1990Q1-2014Q4)*



Note: Real house prices were calculated using data concerning the CPI for Belgium that was retrieved from the website of *Statistics Belgium*. Similarly, we gathered data concerning the CPI in the Netherlands from the website of *Statistics Netherlands*. For both countries the CPI was normalized to 1 for the first quarter of 1985. The data were not seasonally adjusted. The percentage of people with the Dutch nationality living in Belgium and vice versa were provided by *Statistics Belgium* and *Statistics Netherlands*, respectively.

Figure 1 shows that both the evolution of real house prices and the number of

[5]Data from the *National Bank of Belgium* suggests that approximately 21% or 52 billion Euros (12% of 28 billion Euros) of all Belgian imports (exports) comes from (goes to) the Netherlands.

[6]Nederlandse Vereniging van Makelaars, *Dutch Association of Real Estate Brokers*

transactions differed considerably between both countries in recent decades. In a recent paper Damen *et al.* (2016) analyse the evolution of (nominal) house prices across different European countries and provide convincing evidence that these exhibit a long-run relationship with the *ability to pay* of households, which they define as (p. 3) "*[...] a constant fraction of income that goes to housing payments, which results in an amount that people are able to pay based on the possibility to deduct mortgage interest payments and innovative mortgage products.*" The results from their paper suggest that the higher house prices in the Netherlands relative to Belgium are the result of a more generous fiscal regime for owner-occupiers in the former. Dutch households could (and still can) deduct all interest payments from their mortgage loans from their taxable income. The articles by Rouwendal (2007) and Damen *et al.* (2016) also help explaining the observed run-ups observed in the Netherlands since the mid-1990s and in Belgium since 2005. Both of these are likely to be the result of changes in underlying fundamental values. While in the Netherlands the share of interest-only mortgages increased significantly from the 1990s onwards (Rouwendal, 2007), in Belgium the implementation of a more generous fiscal regime in 2005 combined with a (consequent) lengthening in the mortgage term (Damen *et al.*, 2016) increased the *ability to pay*, which subsequently translated into housing prices. It is obvious that these demand shocks, however, are only capitalized into house prices whenever housing supply is relatively inelastic. Vermeulen & Rouwendal (2007) for the Netherlands and Helgers & Buyst (2014) for Belgium, however, show that this is the case for both countries.

After the run-ups in both countries, though, Dutch housing prices and transaction volumes declined after 2008, like in many other (European) countries as a result of the global financial crisis. The average (real) house price and the number of transactions in the Netherlands decreased with respectively 41% and 12% between the first quarters of 2008 and 2009. While the Dutch housing market experienced a bust, the effect of the financial crisis on Belgian housing prices and transaction volumes was only limited. Over the same period (real) house prices and transaction volumes in Belgium decreased only with 1.5% and 14%, respectively. Moreover, transaction volumes in Belgium quickly recovered and even peaked in 2011, while transaction volumes in the Netherlands remain relatively low until the present day. In a recent paper Struyven (2015) attributes the low transaction volumes in the Netherlands to the "housing lock hypothesis". He observes that households who bought their house at the peak have higher Loan-To-Value (LTV) ratios than earlier buyers, and also have much lower mobility rates in every year after purchase.

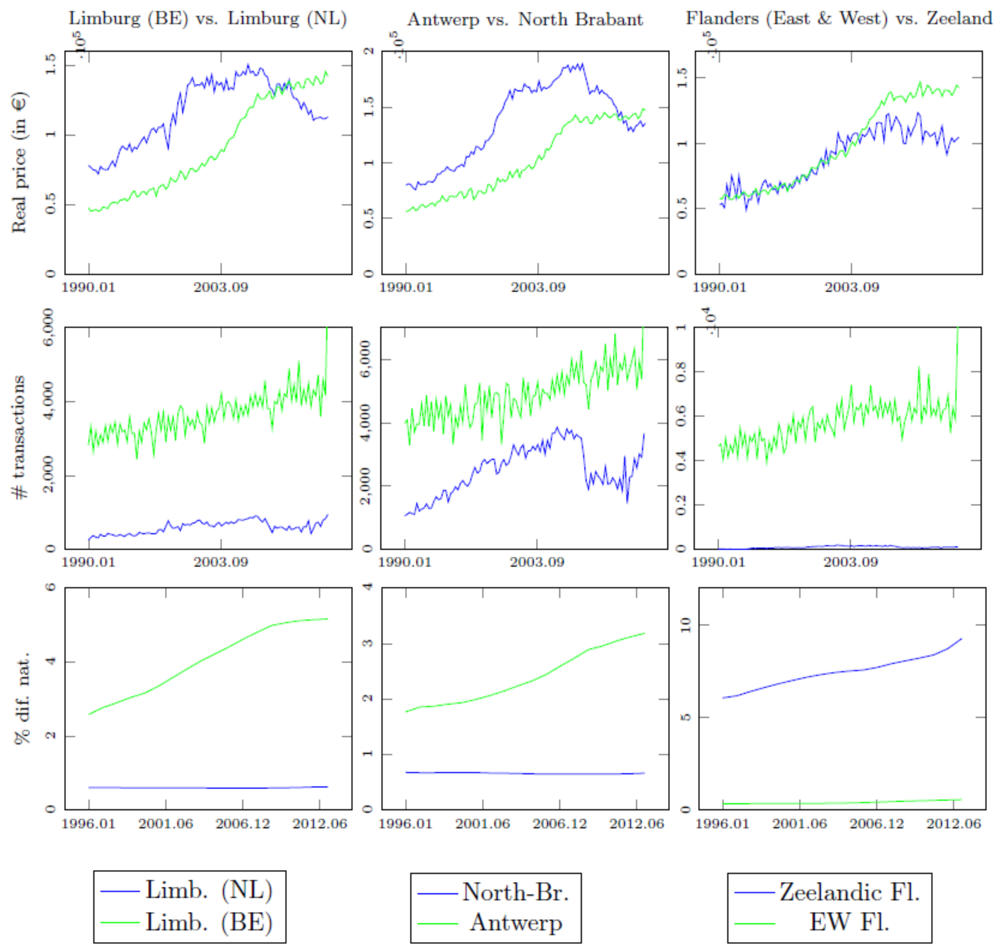
3.2 Developments in border regions

The previous section already showed that developments at the national level in the number of transactions and housing prices have differed considerably between Belgium and the Netherlands in recent decades. These developments at the national level, however, conceal considerable regional differences. It is therefore appropriate to focus on the evolution of housing prices and transaction volumes in regions along the Belgian-Dutch border.

The Belgian-Dutch border spans over a length of approximately 460 kilometers from the North Sea in the west to the tripoint between Belgium, the Netherlands and Germany in the south-east. In the west, the border separates the Belgian provinces East- and West Flanders (Ghent and Bruges) from the sparsely populated Dutch region Zeelandic Flanders (Terneuzen), that is separated from the rest of the province Zeeland (and the Netherlands) by the Western Scheldt which connects the port of Antwerp to the North Sea. The central part of the border separates the Belgian province Antwerp (Antwerp and Turnhout) from the Dutch province North Brabant (Breda, Tilburg and Eindhoven). In the east, the border splits the Belgian provinces Limburg (Tongeren and Genk) and Liège (Liège and Verviers) from the Dutch provinces Noord-Brabant (Eindhoven) and Limburg (Weert, Roermond and Maastricht). Although approximately 20 kilometers of the Belgian-Dutch border separates the French-speaking province Liège from the Dutch-speaking Netherlands, Dutch is the common language spoken on both sides of the border for the remaining 440 kilometers. A graphical overview of the Belgian-Dutch border region is provided in figure A.2 in appendix A.

While house prices might differ because of differences in housing attributes, amenities, etc. spillovers in prices between both countries might also be observed. Since the 2000s Dutch households living in neighboring countries can opt for the Dutch income tax system, which implies that they can benefit from the mortgage interest deductability in the Netherlands. In figure 1 we plot the evolution of real house prices, the number of transactions, and the percentage of people who possess the nationality of the neighboring country for the different Belgian and Dutch provinces that are located along the common border.

Figure 2: *Real house prices, # of transactions, and the % of people with the other nationality for (parts of) the provinces located along the common border (1990Q1-2014Q4)*

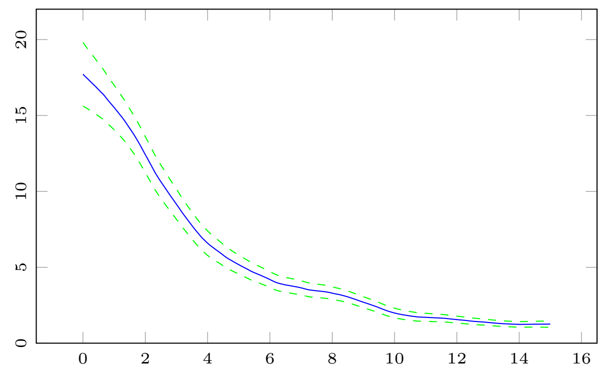


Note: in order to construct average (real) house prices and transaction volumes for the Netherlands we used aggregated data provided by the NVM at the level of 76 designated NVM regions. We then aggregated the data at the level of the provinces. For the province Limburg (NL) we used data for the regions *Zuid-Limburg*, *Roermond eo* and *Weert eo*. The data for the province North Brabant were constructed using data for the regions *Eindhoven eo*, *Zuid Oost Brabant*, *Tilburg Oirschot*, *Breda*, *West Brabant*, and *Bergen op Zoom eo*. For Zeeland we simply took the region Zeelandic Flanders. The % of people with the Belgian nationality for every municipality was provided by *Statistics Netherlands*. We assigned the appropriate NVM region to every municipality using a spatial join procedure in *Quantum GIS*. For Belgium, we used data concerning average house prices, the number of transactions and the % of people with the Dutch nationality at the municipal level and only withheld those municipalities for which the centroid is located within a 15 kilometer radius of the Belgian-Dutch border. The data were then simply aggregated at the provincial level, where we considered the provinces East Flanders and West Flanders to be a single province.

From figure 2 we immediately notice that the evolution of (real) housing prices for the different provinces follow their respective national averages. Note, however, that there are large differences in prices and transaction volumes between the different provinces. While house prices in North Brabant were approximately equal to €180,000 in 2006, those in Zeelandic Flanders were only as high as €120,000. A similar story applies in Belgium. While house prices in Antwerp were approximately equal to €150,000 in 2006, those in Limburg were only as high as €100,000. Furthermore observe that while house prices are generally higher in the Dutch provinces, house prices in East-and West Flanders are higher than in the neighboring Zeelandic Flanders. This can partially be explained by the fact that Zeelandic Flanders is a highly peripheral and sparsely populated region of the Netherlands that lies south of the *Western Scheldt* which separates it from the rest of Zeeland and the country. The (north of the) Belgian provinces East- and West Flanders on the other hand are home to (major) cities such as Ghent and Bruges and are densily populated. Where the higher house prices in the Dutch provinces Limburg and North Brabant have led to an increase in the percentage of people who possess the Dutch nationality living in Belgium, the reverse pattern in the western part of both countries has led to a migration flow from Belgium to the Netherlands. Observe that many people who migrate between both countries remain close to the border of their country of origin, since the percentages of people who possess the nationality of the neighboring country reported in figure 2 are high with respect to their national averages reported in figure 1. A simple kernel-weighted local polynomial of the percentage of inhabitants with the Dutch nationality at the level of the statistical sector[7] versus the distance to the border further confirms this finding.

[7]The statistical sector is the most detailed territorial level Statistics Belgium uses for its statistics and publications. The average statistical sector in Belgium has a territory of about 1.5 square kilometers and houses 700 inhabitants.

Figure 3: *Kernel-weighted local polynomial smoothing: % inhabitants with Dutch nationality vs. distance to border*



Note: the data concerning the number and percentage of inhabitants with the Dutch nationality are aggregated at the level of the statistical sectors in Belgium and were provided by *Statistics Belgium*. The statistical sector is the territorial base unit resulting from a division of the municipalities and the former municipalities by the former National Institute of Statistics for the dissemination of statistics on a more detailed level than the municipal level.

The preference of Dutch households to live close to the Netherlands might have had non-negligible effects on the demand for and prices of Belgian houses close to the border, which is empirically tested in section 6. The evolutions presented in figure 2 also suggest that there is considerable heterogeneity across regions. Heterogeneity of the arbitrage effect is empirically tested in section 7.3. The higher house prices in the Netherlands finally might lead to an *anchoring bias*, as for example suggested by Lambson *et al.* (2004), and may cause Dutch buyers to pay more, *ceteris paribus*, than their Belgian counterparts. This is empirically tested in section 7.4.

4 Data

The main dataset used in subsequent empirical analyses provides detailed information on the sales price and the property and neighborhood characteristics of approximately 26.200 dwellings sold in Flanders between 2003 and 2015 by the agents of a large (Belgian) franchise system of real estate agents. The dataset contains information on the price- and date of sale, interior space (in m^2), plot size (in m^2), number of bedrooms and garages, and so on.[8] Besides a detailed description of the characteristics for every dwelling, we also know its exact coordinates. This allows us to calculate the distance to the border and the distance to other amenities, such as bus stops, highway

[8] A detailed overview of the data is presented in appendix B.

entry/exits, grocery stores and city centers using a Geographic Information System[9].

Table 1: *Descriptive statistics transaction data by region*

Variable	Dist. to border ≤ 15km Obs.	Avg. (1)	St. Dev.	Dist. to border > 15km Obs.	Avg. (2)	St. Dev.	T-test (1)-(2)
Sales price (in €)	5,739	228,231.99	96,409.87	19,983	219,828.50	99,946.11	-5.658***
Living area (in sq. m.)	5,809	183.47	69.61	20,407	184.45	71.91	0.923
Plot size (in sq. m.)	5,809	789.74	1,023.80	20,407	733.12	1,055.15	-3.632***
# Bedrooms	5,808	3.14	0.86	20,402	3.11	0.93	-2.330**
# Garages	5,800	0.84	0.70	20,354	0.83	0.72	-1.527
Year of constr.	5,159	1970.19	22.89	17,490	1965.15	23.09	-13.781***
Year of sale	5,809	2010.34	3.46	20,407	2010.05	3.51	-5.518***
Central heating	5,728	0.79	0.40	20,206	0.70	0.45	-13.255***

The descriptive statistics presented in table 1 reveal that approximately 22% of the houses in the dataset are located less than 15 kilometers from the Belgian-Dutch border. Although these 5,800 properties will thus be of special interest in the subsequent empirical analyses, we will use samples of the remaining 20,400 transactions to estimate several placebo-tests in section 7.2. The descriptive statistics show that the dwellings located close to the border are, on average, more expensive than their inland counterparts (228,000 vs. 220,000). Although this may seem counterintuitive at first, it has to be noted that some of the larger and more expensive Flemish cities (e.g. Bruges, Ghent and Antwerp) are located close the Belgian-Dutch border. The descriptive statistics furthermore reveal that the average dwelling within 15 kilometers of the border is located on a larger plot ($789m^2$ vs. $733m^2$), has more bedrooms (3.14 vs. 3.11) and was constructed more recently (1970 vs. 1965). The detailed transaction data presented above allow us to estimate (spatial) hedonic models. The estimation strategy is presented in the next section.

5 Methodology

Valuing a home has traditionally proven to be difficult. Every home has its own specific location and its own (unique) set of characteristics that may affect its value. Nonetheless, a large body of literature has attempted to explain the value of housing by valuing its individual components using the

[9]A Geographic Information System is a system designed to capture, store, manipulate, analyse, manage, and present all types of spatial or geographical data

hedonic pricing model.[10] The hedonic pricing model (HPM), for which the theoretical foundations date back to the seminal works by Lancaster (1966) and Rosen (1974), departs from the idea that utility is not generated by the (composite) good per se, but by the characteristics that define the good. In his excellent overview of the literature Malpezzi (2003) illustrates this idea perfectly by stating: "*I'm happy to be home, not so much to be in anything called a 'house', so much as to be in a warm dry place, with a quiet space for a comfortable chair, a functioning toilet or a hot bath should I require them, and some other rooms in the house to store stacks of papers or noisy children.*" Following Rosen (1974), hedonic prices are defined as "*the implicit prices of attributes and are revealed to economic agents from observed prices of differentiated products and the specific amounts of characteristics associated with them.*" Traditionally, the HPM is estimated using ordinary least squares (OLS) where the estimated coefficients represent the shadow prices of the attributes. In the current paper we allow for a more flexible form and estimate variants of the following regression equation:

$$\ln(p) = \underbrace{\beta X}_{\text{Classical hedonic pricing model}} + \overbrace{\gamma L}^{\text{Spatial arbitrage}} + \underbrace{\lambda \ln(p)}_{\text{Spatial lag component}} + u$$
$$\text{where}$$
$$u = \underbrace{\rho W u + \epsilon}_{\text{Spatial error component}} \tag{II.1}$$

where $\ln(p)$ denotes a vector of logarithmically transformed sales prices, X a matrix of the individual houses' structural and neighborhood characteristics as described in section 4, and β the vector of corresponding implicit prices. The matrix L represents location-specific information, which consists of two components. Firstly, we control for the proximity to centers of employment (e.g. Antwerp, Ghent), various amenities and the municipality the property is located in. Since the Flemish-Dutch border stretches out over approximately 460 kilometer, house prices might differ substantially across locations. Secondly, and more interestingly for the current paper, we augment the regression with distance to the border in polynomial form and interaction-effects between distance to the border and other variables.

[10] Excellent overviews of the use of the hedonic pricing model in housing economics can be found in Malpezzi (2003) and Sirmans *et al.* (2005).

While the regression analysis presented so far could be carried out using ordinary least squares (OLS), the regression coefficients could be biased due to spatial autocorrelation in the dependent and independent variables. We therefore potentially allow for spatial autocorrelation in the dependent variable (spatial lag component), $\ln(p)$, and spatial autocorrelation in the residuals (spatial error component), u. The spatial structure of the model is captured by the spatial weights matrix W, where the elements w_{ij} represent the spatial dependency between observation i and observation j and are equal to zero whenever $i = j$.[11] More specifically, we construct a row-normalized[12] *10 nearest neighbors* and a row-normalized *inverse-distance* spatial weights matrix. The models are estimated using the *Generalized Spatial 2 Stage Least Squares* (GS2SLS) estimator in Stata 14 that was proposed by Kelejian & Prucha (1998).[13]

6 Baseline results

In this section we report and discuss the results for our baseline analyses. In table C.3 in appendix C the estimated coefficients of a classical (spatial) hedonic model are presented. The coefficients for the house-specific (e.g. structure size, plot size, # bedrooms,..) and neighborhood-specific (e.g. population density, distance to city hall) are familiar and, in general, have the expected signs and are statistically significant. Therefore, we proceed by estimating a number of (spatial) hedonic pricing models where we investigate whether houses located closer to the Flemish-Dutch border are more expensive. We also examine temporal and spatial heterogeneity by allowing the *distance to the border* variable to interact with a time trend and an urban area dummy variable. The results of our baseline analysis are presented in table 2.

[11] An observation cannot be dependent upon itself.

[12] This is conventional in the spatial econometrics literature.

[13] The models could also be estimated using a Maximum Likelihood estimator, but the GS2SLS estimator additionally allows for heteroskedasticity and is superior in terms of speed. Excellent discussions of the generalized method of moments and instrumental variables estimation approach underlying the GS2SLS-estimator can be found in Arraiz *et al.* (2010) and Drukker *et al.* (2013).

Table 2: *Estimation results distance to the border (spatial error model, 10 nearest neighbors spatial weights matrix)*

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Dist. to border	-0.0031	-0.0085***	-0.0071	-0.0222**	-0.0247***	-0.0316***
	(0.0021)	(0.0027)	(0.0047)	(0.0087)	(0.0088)	(0.0094)
Dist. to border^2			-0.0001	0.0026*	0.0026**	0.0021
			(0.0003)	(0.0013)	(0.0013)	(0.0013)
Dist. to border^3				-0.0001**	-0.0001**	-0.0001**
				(0.0001)	(0.0001)	(0.0001)
Dist. to border*(Year of sale-2003)		0.0006***	0.0006***	0.0006***	0.0006***	0.0006***
		(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Dist. to border*Urban area					0.0062*	0.0054
					(0.0037)	(0.0037)
Dist. to border*Travel time Dutch city						0.0005**
						(0.0002)
ρ	0.2032***	0.2045***	0.2045***	0.1976***	0.197***	0.1961***
	(0.0325)	(0.0326)	(0.0326)	(0.033)	(0.033)	(0.0331)
Obs.	4,657	4,657	4,657	4,657	4,657	4,657
R-sq.	0.833	0.833	0.833	0.833	0.833	0.834

Note: ***, ** and * denote that the estimated regression coefficients are statistically significant at the 1, 5 and/or 10 percent level. In all the regression models presented we control for an extensive list of property- and neighborhood characteristics equivalent to those presented in table C.3 in appendix C.

The results presented in table 2 generally suggest that dwellings located closer to the border are more expensive. In the simple specification presented in column (2), where we include the distance to the border linearly and allow the distance to the border to interact with a simple time trend, we find that dwellings located 15 kilometers from the Flemish-Dutch border were, *ceteris paribus*, approximately 13% cheaper in 2003 than their equivalent counterparts located "at the border". The quadratic and cubic specifications reported in columns (3) and (4) show very similar results where comparable houses that are 15 kilometers from the border were 13 and 16 percent less expensive in 2003 as can be observed from figure 4. Our estimates thus confirm that the increase in demand for Flemish dwellings close to the Flemish-Dutch border led to substantially higher prices along the border.

Notice that the distance to the border only becomes statistically significant when it is also allowed to interact with the year of sale.[14]. The reported coefficients in columns (2)-(6) suggest that the distance to the border effect decreases (in absolute value) with 0.06 percent per kilometer per year. The estimates reported in column (2), for example, suggest that while houses located 15 kilometers from the Flemish-Dutch border were approximately 13 percent cheaper in 2003, this effect was almost negligible (1.95%) in 2015.[15] This finding is consistent with the boom in Belgian housing markets since 2005 and the subsequent collapse of the Dutch housing market in the aftermath of the global financial crisis of 2007-2008. While Belgian property prices generally rose due to a more generous fiscal treatment of owner-occupiers (Damen *et al.*, 2016), the (Dutch) demand for properties close to the border decreased. As a result, the prices of properties close to and further from the border equalized and the distance to the border has become less important over time.

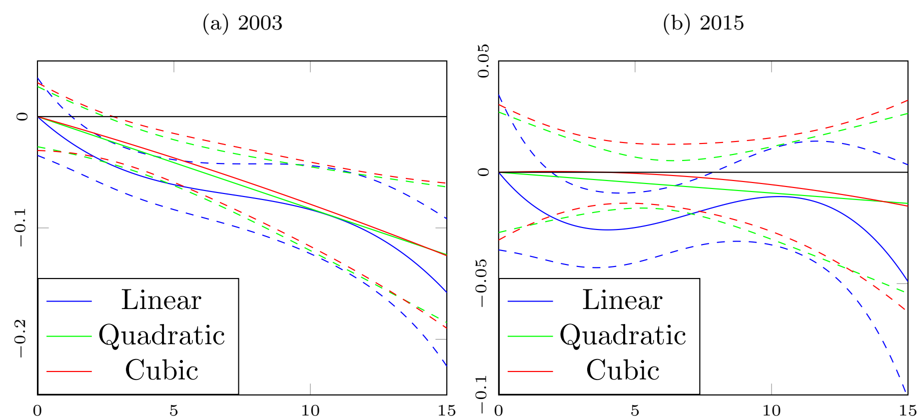
In columns (5) and (6) we furthermore include interaction effects between the distance to the border and (1) a dummy-variable that is equal to one whenever the property is located in an urban area, and (2) the travel time

[14]Notice that we normalized the year of sale, such that the estimated coefficient for distance to the border variable represents the arbitrage effect in 2003, the first year for which we have data.

[15]$-0.0085^{***} * 15 + 0.0006^{***} * 12 * 15 \approx -0.0195$

to the nearest Dutch city. [16] [17] The estimates reported in column (6) indicate that the distance to the border effect is smaller, albeit not statistically significant, in urban areas and larger when the travel time to the nearest Dutch city is lower. An interpretation for the first effect is that property prices in (Flemish) urban areas are generally higher, which limits the scope for arbitrage. The reported positive coefficient (0.0005) for the second interaction effect suggests that the demand for close to the border is higher whenever the travel time to the nearest Dutch city is lower, which is also intuitive. All in all, the results presented suggest that the distance to the border effect is not constant across time and space.[18]

Figure 4: *Visualization estimated effects distance to the border*



Note: the estimated effects and confidence bounds for columns (2), (3) and (4) of table 2 are depicted in the figure for the years 2003 and 2015.

7 Robustness & extensions

Although the results presented in section 6 already shed light on the relationship between housing prices and proximity to the border, it remains to perform a battery of robustness tests and extensions. In appendix C we show that the reported coefficients are robust with respect to the type of

[16] To determine whether a property is located in an urban area, we use the classification of Luyten & Van Hecke (2007) who allocate the 589 Belgian municipalities into one of 18 urban districts or the remaining rural area.

[17] We therefore used the geo-coordinates of the following Dutch cities (from west to east): *Terneuzen*, *Bergen op Zoom*, *Roosendaal*, *Breda*, *Tilburg*, *Eindhoven*, *Weert*, *Sittard-Geleen* and *Maastricht*.

[18] In section 7 we allow more generally for spatial heterogeneity.

spatial spillovers (OLS[19] and spatial lag model[20]) and the specification of the spatial weights matrix (inverse-distance[21]). In section 7.1, we estimate the same model, but use the travel time to the nearest border crossing as our variable of interest. In section 7.2, we carry out several placebo tests. The results presented in these sections, if anything, make our earlier findings more convincingly. In section 7.3, we assess whether the estimated spatial arbitrage effect is homogeneous across space and in section 7.4 we investigate whether Dutch buyers pay a premium for comparable housing compared to their Belgian counterparts.

7.1 Travel time to border crossing

In the previous section we used the euclidean distance between the transaction and the border as our spatial arbitrage measure. It might however be the case that, for example, natural barriers that coincide with the border hamper arbitrage between regions on both sides of the border.[22] Therefore, we have also calculated the minimum travel time between every transaction and the nearest border crossing.[23] We now re-estimate the same model but use the alternative distance measure. The findings are reported in table 3.[24]

[19]Table C.4.

[20]Table C.5.

[21]Tables C.6 and C.7.

[22]This is especially likely to be the case along the eastern border of Flanders where the river *Meuse* separates Flanders from the Netherlands.

[23]In a first step, we identified the geographical coordinates of 76 border crossing using a Geographical Information System and OpenStreetMap data. In a second step, we used the geographical coordinates of properties and border crossings, respectively, to calculate all mutual travel times ($\approx 354,000$ combinations) using the *osrmtime*-algorithm (Huber & Rust, 2016) available in Stata 14. In a final step we only kept the travel time to the nearest border crossing for every transaction in our dataset.

[24]Similar robustness checks to those presented for the baseline model are presented in tables C.8, C.10, C.9 and C.11 in appendix C.

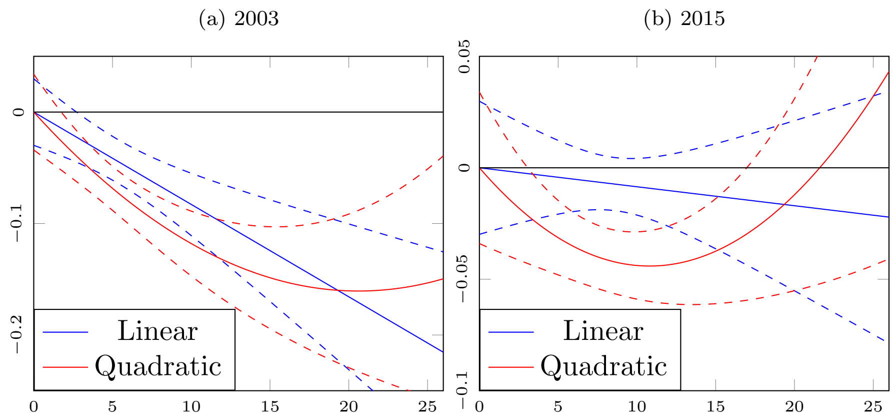
Table 3: *Estimation results travel time to the nearest border crossing (spatial error model, 10 nearest neighbors spatial weights matrix)*

Variable	(1)	(2)	(3)	(4)	(5)
Travel time border	-0.0031*	-0.0083***	-0.0156***	-0.016***	-0.0158***
	(0.0017)	(0.0021)	(0.0034)	(0.0036)	(0.0042)
Travel time border^2			0.0004***	0.0004**	0.0004
			(0.0001)	(0.0001)	(0.0002)
Tr. time border*(Year of sale-2003)		0.0006***	0.0006***	0.0006***	0.0006***
		(0.0002)	(0.0002)	(0.0002)	(0.0002)
Tr. time border*Urban area				0.0012	0.0012
				(0.0026)	(0.0026)
Tr. time border*Tr. time Dutch city					0
					(0.0002)
ρ	0.2047***	0.2086***	0.2075***	0.2069***	0.207***
	(0.0325)	(0.0326)	(0.0326)	(0.0326)	(0.0326)
Obs.	4,657	4,657	4,657	4,657	4,657
R-sq.	0.834	0.835	0.835	0.835	0.835

Note: ***, ** and * denote that the estimated regression coefficients are statistically significant at the 1, 5 and/or 10 percent level. In all the regression models presented we control for an extensive list of property- and neighborhood characteristics equivalent to those presented in table C.3 in appendix C.

The results reported in table 3 are similar to those presented in table 2, as is to be expected since both measures are highly correlated. The reported coefficient in column (2) for the travel time to the border (-0.0083), for example, indicates that a 10 minute increase in travel time, *ceteris paribus*, is associated with a 8.3% decline in housing prices. Again note that this effect is decreasing over time. Where a 10 minute increase in travel time to the border lowers home values with 8.3% in 2003, this effect reduces to 1.1% in 2015. This is also consistent with our previous findings. In columns (3)-(5) we once more allow for non-linearities and interaction effects. The results suggest that the marginal effect of travel time to the border on housing prices is strong for low travel time, but slowly decreases. Also observe that we no longer find any evidence supporting interaction effects. The results of the regression analyses are graphically portrayed in figure 5.

Figure 5: *Visualization estimated effects travel time to border crossing*



Note: the estimated effects and confidence bounds for columns (2), (3) and (4) of table 3 are depicted in the figure for the years 2003 and 2015.

7.2 Placebo-testing

While the results presented in tables 2 and 3 suggest that dwellings that are/were located in closer proximity to the border are more expensive due to spatial arbitrage, it might be the case that the distance to the border is correlated with other unobservables. As a result, the estimated spatial arbitrage effect is potentially biased. To assess the robustness of the results presented in the previous sections, we therefore perform several placebo tests. Although placebo testing has become the standard practice in for example medical sciences, its use in econometrics has been far less common. Fortunately, our dataset does not only contain information on transactions that are located within 15 kilometers of the actual border, but also contains information on transactions located in other parts of Belgium. It is therefore relatively straightforward to carry out several placebo tests. More specifically, we perform these tests by "shifting" the Belgian-Dutch border and the corresponding Belgian territory southwards by 5, 10, 15, 20, 30 and 40 kilometers, respectively.[25] Subsequently, we calculate the distance to the new border for all transactions located in the new hypothetical Belgium and only retain those transactions located within 15 kilometers from the new placebo border. We then simply repeat the analyses presented previously. The results of the placebo tests are presented in table 4.

[25]In figure C.4 in appendix C the newly created hypothetical borders are presented. Similar to the baseline analysis, we only include properties in the regression analysis that are located within 15 kilometers of the placebo border.

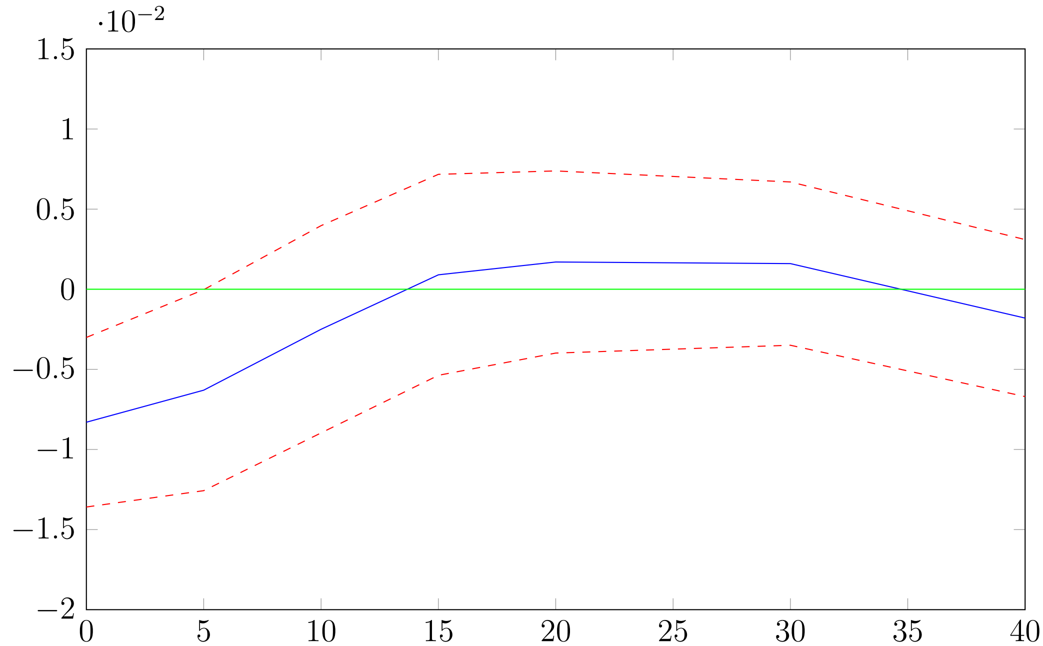
Table 4: *Estimated placebo-effects*

Placebo border	0km	5km	10km
Dist. to border	-0.0083***	-0.0063*	-0.0025
	(0.0027)	(0.0032)	(0.0033)
Dist. to border*(Year of sale-2003)	0.0006***	0.0011***	0.0008***
	(0.0002)	(0.0002)	(0.0002)
ρ	0.2054***	0.3499***	0.3591***
	(0.0326)	(0.0271)	(0.0284)
Obs.	4,657	4,453	4,178
R-sq.	0.834	0.814	0.827
Placebo border	20km	30km	40km
Dist. to border	0.0017	0.0016	-0.0018
	(0.0029)	(0.0026)	(0.0025)
Dist. to border*(Year of sale-2003)	-0.0003	-0.0003	0
	(0.0002)	(0.0002)	(0.0002)
ρ	0.3204***	0.1934***	0.2665***
	(0.0317)	(0.0317)	(0.0244)
Obs.	3,886	4,585	6,432
R-sq.	0.823	0.809	0.826

Note: ***, ** and * denote that the estimated regression coefficients are statistically significant at the 1, 5 and/or 10 percent level. In all the regression models presented we control for an extensive list of property- and neighborhood characteristics equivalent to those presented in table C.3 in appendix C.

The results from the placebo tests support the hypothesis that the reported coefficients for the distance to the border are the result of spatial arbitrage, since the estimated coefficients for our variable of interest are close to zero and no longer statistically significant for all but one of the placebo borders. Note that the only placebo border for which we find a statistically significant negative effect is 5 kilometers, which is intuitive since the placebo border is still in close proximity to the actual border. Furthermore note that the estimated effect for the 5 kilometer placebo border is smaller than that of the original border, which is consistent with the earlier finding that the initial effect of moving away from the border is strong but then slowly decays. The results are also graphically portrayed in figure 6.

Figure 6: *Estimated spatial arbitrage effects placebo borders*



Note: the blue line displays the point estimates for various placebo tests. The dashed red lines represent the corresponding confidence bounds at the 95% level.

7.3 Spatial heterogeneity

Since the Flemish-Dutch border stretches out over 460 kilometers and covers multiple housing markets, it is likely that the estimated arbitrage effect is not homogeneous across space. Although methods, such as *Geographically Weighted Regression* (GWR; Brunsdon *et al.*, 1996; Fotheringham *et al.*, 2003), have been developed to model spatially heterogeneous processes, we employ a much simpler estimation strategy in the current paper. More specifically, we divide the border into segments of predefined lengths and calculate the distance to the closest border segment for every transaction in our dataset. Since choosing the length of the border segments is arbitrary, we conduct the analysis for various lengths (5, 10, 20, 30 and 40 kilometers) of the border segments. An overview of the results is presented in table 5.

Table 5: *Overview spatial heterogeneity border effect*

Length border segments	Neg.	Neg. & stat. sign.	Pos.	Pos. & stat. sign.	# Border segments
5km	34	10	11	1	45
	75.56%	22.22%	24.44%	2.22%	
10km	24	6	5	0	29
	82.76%	20.69%	17.24%	0%	
20km	18	6	1	0	19
	94.74%	31.58%	5.26%	0%	
30km	14	6	0	0	14
	100%	42.86%	0%	0%	
40km	11	6	0	0	11
	100%	54.55%	0%	0%	

Note: in table C.12 in appendix C.2 all the estimated slope coefficients and their respective standard errors are presented. Kernel density plots and maps of the estimated coefficients are also presented in figures C.4 and C.5 in appendix C.2.

When we divide the border into segments of 5 kilometer, our most flexible specification, our estimates reveal that the slope coefficient is negative for 34 out of the 45 border segments and negative and statistically significant for 10 regions. Observe furthermore that the estimated slope coefficient is positive and statistically significant for only one border segment. A thorough examination of a map of the estimation results presented in figure C.5 in appendix C reveals that the estimated arbitrage effects are especially strong in more rural regions that are mostly located east of Antwerp. These findings are intuitive, since the scope for arbitrage was large in these regions as relatively inexpensive Belgian regions are neighboring more expensive Dutch regions there.

7.4 Do Dutch Buyers Pay a Premium?

While the traditional hedonic pricing model has developed into one of the main workhorses of housing economists who investigate the determinants of housing prices, more recently researchers have questioned some of the underlying hypotheses. The traditional hedonic model views a heterogeneous good as a collection of characteristics, each of which has a well-defined shadow price. Markets are furthermore assumed to be sufficiently thick, such that the shadow prices of characteristics are revealed to all agents. As was argued by Harding *et al.* (2003), these assumptions imply that there is no explicit role for bargaining in the traditional hedonic model. In a number of related recent studies, however, Lambson *et al.* (2004) and Ihlanfeldt & Mayock (2012) find that out-of-state buyers pay significant premiums for (observationally) equivalent dwellings. Lambson *et al.* (2004) argue and also find evidence which suggests that these premia are likely to be the result of search costs, biased beliefs (anchoring bias) and haste associated with out-of-state buyers.

In the current setting we are fortunate enough that the real estate agencies reported the previous address of 9,720 buyers in our full sample ($\approx$ 37%).[26] We observe that the previous address is located in the Netherlands for 183 buyers ($\approx$ 1.9%). The data furthermore reveals that there are 1,989 transactions located less than 15 kilometers from the border for which the address is given. For this subset of transactions, there are 153 buyers ($\approx$ 7.7%) whose previous address was located in the Netherlands, which is consistent with the earlier findings in figure 3 that Dutch buyers prefer to live close to the border. A further exploration of the data suggests that the Dutch buyers in our sample cluster spatially. 144 of the 153 Dutch buyers in our sample bought a property that is located in the adjacent districts *Maaseik* and *Turnhout*, where their share of the market is about 11%. Given this clustering of Dutch buyers and the findings presented in the previous section, we limit the analysis to the 1,337 transactions in these two districts. In table 6 we present some descriptive statistics for the dwellings bought by Dutch and Belgian buyers, respectively.

Table 6: *Descriptive statistics transaction data by origin buyer*

		Buyer = Dutch			Buyer = Belgian		T-test
Variable	Obs.	Avg. (1)	St. Dev.	Obs.	Avg. (2)	St. Dev.	(1)-(2)
Sales price (€)	144	260,554.30	88,638.13	1,190	224,783.24	84,886.97	4.753***
Living area (sq. m.)	144	217.25	64.87	1,193	180.58	66.01	6.308***
Plot size (sq. m.)	144	1,128.37	837.62	1,193	805.56	1,024.56	3.636***
# Bedrooms	144	3.42	0.87	1,193	3.13	0.82	3.949***
# Garages	144	1.04	0.62	1,192	0.80	0.64	4.171***
Year of constr.	137	1976.15	17.72	1,104	1972.10	23.68	1.935*
Year of sale	144	2007.84	2.23	1,193	2010.01	2.82	-8.908***
Central heating	144	0.96	0.18	1,191	0.86	0.34	3.543***
Dist. border (km)	144	4,464.94	3,416.87	1,193	6,072.69	3,453.37	-5.283***
Dist. buyer - pr. (km)	144	44.71	40.59	1,193	9.03	16.69	19.606***

The descriptive statistics presented indicate that, on average, Dutch buy-

[26]The total dataset contains 26,216 transactions.

ers buy dwellings that are not only more expensive but also differ in their structural characteristics and are located closer to the border. Observe furthermore that the average year of sale of Dutch buyers is lower, which is consistent with a decrease in the demand of Dutch buyers due to a bust in the Dutch housing market after 2008. The results of various (spatial) hedonic models are presented in table 7.

Table 7: *Do Dutch buyers pay a premium*

Variable	(1)	(2)	(3)	(4)
Buyer NL	0.2795***	0.159***		0.1178**
	(0.0622)	(0.045)		(0.0475)
Buyer NL*(Year of sale - 2003)	-0.0336***	-0.0182**		-0.0144*
	(0.0113)	(0.0065)		(0.0066)
Buyer NL*Dist. to border		-0.0099*		-0.0057
		(0.0044)		(0.0046)
Buyer NL*Min. travel time				
Min. travel time	0.0046		-0.0053	-0.0055
	(0.0048)		(0.0032)	(0.0032)
Dist. to border	-0.0317***		-0.0188***	-0.0148**
	(0.0087)		(0.0056)	(0.0059)
Dist. to border*(Year of sale-2003)	0.0026***		0.0017***	0.0012*
	(0.0005)		(0.0005)	(0.0006)
ρ	0.7481***	0.2398**	0.2118*	0.216*
	(0.0444)	(0.1028)	(0.1055)	(0.1062)
Observations	1,271	1,155	1,155	1,155
R-sq.	0.072	0.831	0.831	0.832
Property char.	No	Yes	Yes	Yes
Year of sale	No	Yes	Yes	Yes
Municipality dummies	No	Yes	Yes	Yes

Note: ***, ** and * denote that the estimated regression coefficients are statistically significant at the 1, 5 and/or 10 percent level. In all the regression models presented we control for an extensive list of property- and neighborhood characteristics equivalent to those presented in table C.3 in appendix C.

In the first column of table 7 we report the estimated coefficients of an "empty model". While the results suggest that Dutch buyer pay a large premium relative to their Belgian counterparts, we have to note that the model does not take into account differences in the structural characteristics of properties and does not contain any locational information. The estimates presented in the second column indicate that a substantial part ($\approx 12\%$) of the difference in average transaction prices, however, can be explained by differences in the dwellings bought by Dutch and Belgian buyers. Moreover, the results presented in the fourth column suggest that the effect further declines with 0.0412 after controlling for spatial arbitrage effects. The results reported in the fourth column indicate that, in 2003, the premium paid by Dutch buyers was equal to 11.8%. The estimates, however, also indicate that this premium has been steadily declining over time. Since the gap between Belgian and Dutch housing prices has also steadily declined over time, we interpret this finding as support for the *anchoring-bias* hypothesis. Finally note, by comparing the estimated coefficients for the distance to the border variable in the third and fourth column, that the spatial arbitrage effect can partially be explained by Dutch buyers who pay a premium since the estimated value of the distance to the border coefficient declines in absolute value when controlling for the origin of the buyer.

8 Concluding Remarks

In this paper we have investigated the determinants of Flemish housing prices along the Flemish-Dutch border and focused on the *distance to the border effect* using a large sample of detailed individual transaction data provided by a large franchise system of real estate agencies. The results from various (spatial) hedonic models suggest that the price of a dwelling located at the border was about 13% higher in 2003 than an observationally equivalent dwelling located 15 kilometers from the border. This finding is consistent with spatial arbitrage by Dutch households who took advantage of the lower housing prices in Flanders, but preferred to live close to the Netherlands. The results found in various regression analyses also suggest that this effect has been steadily decreasing over time, where the initial gap of 13% in 2003 has decreased to 2% in 2015. This is in accordance with the "catching up" of Flemish housing prices as a result of a more generous fiscal regime for owner-occupiers in Belgium since 2005 and a bust in the Dutch housing market in the aftermath of the global financial crisis of 2007-2008. These (independent) developments have reduced the scope and desirability of arbitration, as shown by the estimates of our hedonic model.

In a series of robustness tests we show that these results are robust with respect to alternative econometric specifications and alternative measures of distance to the border. Since the dataset does not only contain information on transactions close to the border, we furthermore carry out a series of placebo-tests that further strengthen our previous findings. In the last part of the paper we show that the distance to the border effect is especially pronounced in rural areas in the eastern parts of Flanders. Using a subsample of the full dataset for which the previous address of the buyer is known and focusing on a well-defined local market, we finally show that Dutch buyers pay a premium for observationally equivalent dwellings relative to their Belgian counterparts. This premium, however, is decreasing over time which is consistent with the notion of anchoring bias.

Although the paper shows a coherent series of findings, we unfortunately have only detailed data available for real estate transactions on one side of the border. In a next step, it would be interesting to supplement the current dataset with information concerning transactions on the other side of the border, as for example done in Micheli (2014). While these authors study the magnitude of the border effect at the Dutch-German border, two countries that do not share a single common language, combining the current dataset with data concerning Dutch real estate transactions would allow us to study the magnitude of the border effect between countries that share a common language. This would, for example, help us to develop a more thorough understanding of the European integration process.

A The Belgian-Dutch border

Figure A.1: *Graphical representation border effects*

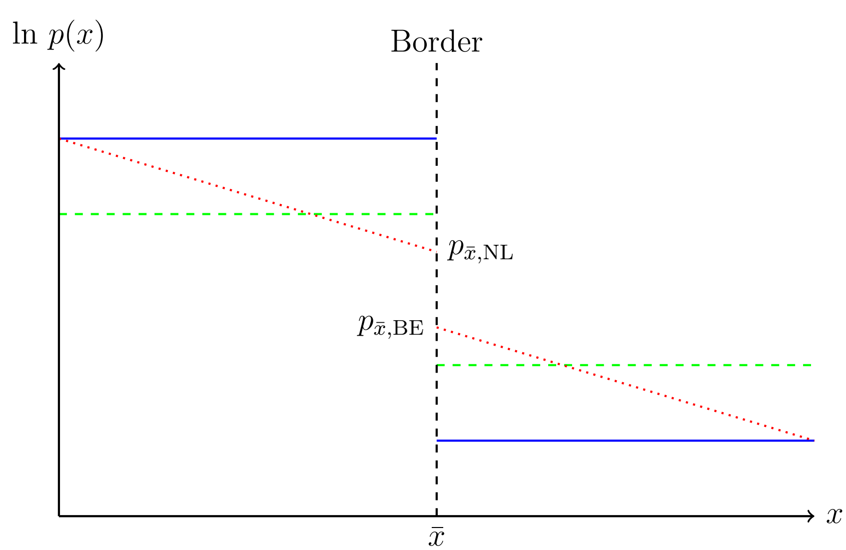
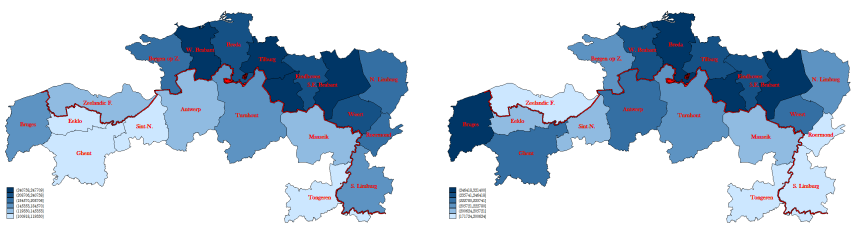


Table A.1: *Moran's I for % Dutch inhabitants*

Spatial Weights Matrix	I	$E[I]$	$sd[I]$	z	p-value
Negative exponential	0.67405	-0.00037	0.00746	90.350	0
Inverse distance	0.15799	-0.00037	0.00130	121.640	0

Figure A.2: *Nominal house prices in different regions (2003 & 2011)*

(a) 2003 (b) 2011



Source maps: Belgian HISGIS & Statistics Netherlands. Source data: Statistics Netherlands & Statistics Belgium.

B Overview of the data

Table B.2: *Variables & sources of data*

Source	Variable	Description
ERA Belgium	Transaction price	Sales price (in euro)
	Date of sale	
	Location	x- and y-coordinates
	Type of construction	(Semi-)detached vs. terraced housing
	Year of construction	
	Living surface (in sq. m.)	
	Plot size (in sq. m.)	
	Number of bedrooms	
	Number of garages	
	State of property	Ready to move in, luxuriously finished
	Heating type	Central heating
	Heating material	Gas, wood, electricity,..
	Heating elements	Radiators, underfloor heating,..
	Hot water	Condensing boiler,..
	Glazing	Single, double, triple,..
	Basement	Storage room, wine cellar,..
	Bathroom	Two/multiple bathrooms, bathtub,..
	Kitchen	Well-maintained, dishwasher,..
	Various	Fireplace, alarm, airconditioning,..
Fl. Geogr. Inf. Agency (AGIV)	Bus stops	x- and y-coordinates
	Train stations	
	Highway entries/exits	
	Grocery stores	
	Shopping centers	
Statistics Belgium (ADS)	Transaction prices	1973Q1-2014Q4, municipalities
	Number of transactions	
	Number of inhabitants	1996-2013, municipalities
Statistics Netherlands (CBS)	Number of inhabitants	1996-2013, municipalities

C Regression results

C.1 Baseline results

Table C.3: *Estimated coefficients control variables*

Category	Variable	$\hat{\beta}_k$ ($\hat{\sigma}_k$)	Category	Variable	$\hat{\beta}_k$ ($\hat{\sigma}_k$)
Building type	Semi-detached	0.0248***	Glazing	Single	-0.0532***
		(0.0076)			(0.006)
	Detached	0.0542***		Double	0.0199***
		(0.0089)			(0.0064)
Size	Ln(living surf.)	0.2449***		Triple	0.0216
		(0.0098)			(0.0234)
	Ln(Plot size)	0.1512***	Kitchen	Dishwasher	0.0453***
		(0.0044)			(0.0056)
Age	Linear	-0.0042***		Well-maintained	0.025***
		(0.0004)			(0.0056)
	Squared	0***		All comfort	0.0249***
		(0)			(0.0058)
# Bedrooms	0	-0.0253	Bathroom	Two bathrooms	0.0214***
		(0.0892)			(0.0079)
	1	-0.1156***		Multiple bathrooms	0.0305*
		(0.0206)			(0.0181)
	3	0.0464***		Ind. bathr. per. bedr.	0.2341***
		(0.0068)			(0.0553)
	4	0.0583***		Double sink	0.0204***
		(0.0084)			(0.006)
	5	0.0797***		Separate toilet	0.0167***
		(0.0129)			(0.005)
	6	0.0701***	Basement	Total basement	-0.0039
		(0.0261)			(0.0107)
	7	0.17***		Wine cellar	0.0615***
		(0.0404)			(0.0201)
# Garages	1	0.0305***	Environment	Residential area (1)	0.0319***
		(0.0062)			(0.0075)
	2	0.0341***		Residential area (2)	0.0637***
		(0.0093)			(0.0125)
	3	0.0384*		Forest/parc	0.0221
		(0.0205)			(0.0141)
	4	0.0167		Unobstructed view	0.0184**
		(0.0393)			(0.0081)
State	Luxuriously finished	0.0868***	Various	Automatic garage door	0.0116*
		(0.0112)			(0.0064)
	Ready to move in	0.0906***		Alarm	0.0723***
		(0.0062)			(0.0101)
	Light refresh necessary	-0.0053	Neighborhood	Dist. to city hall	-0.0086***
		(0.0061)			(0.0018)
	Major refurbishment necessary	-0.0418***		Travel time city	-0.0058***
		(0.0108)			(0.0018)
	Total reconstruction necessary	-0.2448***		Ln(median tax. inc.)	0.0784***
		(0.0217)			(0.0246)
Heating type	Central heating	0.0806***		Ln(pop. density)	0.007**
		(0.0092)			(0.0027)
Heating material	Gas	0.0265***		Travel time Dutch city	-0.0002
		(0.0057)			(0.0013)
	Electricity	-0.0026	Constant	Constant	9.0588***
		(0.0131)			(0.2956)
Heating elements	Radiators	0.021***		ρ	0.2057***
		(0.0081)			(0.0324)
	Underfloor heating	0.0378***	Diagnostics	Obs.	4,657
		(0.0098)			
	Convectors	0.0093		R-sq.	0.828
		(0.0088)			
	Accumulation	0.0216		Municipality dummies	Yes
		(0.0173)			
Hot water	Condensing boiler	0.0219***		Year dummies	Yes
		(0.008)			

Note: ***, ** and * denote that the estimated regression coefficients are statistically significant at the 1, 5 and/or 10 percent level.

C.2 Robustness & extensions

Table C.4: *Estimation results distance to the border (ordinary least squares)*

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Dist. to border	-0.0029	-0.0082***	-0.0069	-0.0238***	-0.0259***	-0.0321***
	(0.002)	(0.0026)	(0.0043)	(0.0084)	(0.0085)	(0.0087)
Dist. to border^2			-0.0001	0.0029**	0.0029**	0.0024*
			(0.0003)	(0.0013)	(0.0013)	(0.0013)
Dist. to border^3				-0.0001**	-0.0001**	-0.0001**
				(0.0001)	(0.0001)	(0.0001)
Dist. to border*(Year of sale-2003)		0.0006***	0.0006***	0.0006***	0.0006***	0.0006***
		(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Dist. to border*Urban area					0.0055	0.0048
					(0.0035)	(0.0035)
Dist. to border*Travel time Dutch city						0.0005**
						(0.0002)
Obs.	4,657	4,657	4,657	4,657	4,657	4,657
R-sq.	0.833	0.833	0.833	0.833	0.833	0.834

Note: ***, ** and * denote that the estimated regression coefficients are statistically significant at the 1, 5 and/or 10 percent level. In all the regression models presented we control for an extensive list of property- and neighborhood characteristics equivalent to those presented in table C.3 in appendix C.

Table C.5: *Estimation results distance to the border (spatial lag model, 10 nearest neighbors spatial weights matrix)*

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Dist. to border	-0.0032*	-0.0086***	-0.0076*	-0.0204***	-0.0214***	-0.0287***
	(0.0018)	(0.0024)	(0.0041)	(0.0075)	(0.0076)	(0.0081)
Dist. to border^2			-0.0001	0.0022*	0.0022*	0.0015
			(0.0002)	(0.0011)	(0.0011)	(0.0012)
Dist. to border^3				-0.0001**	-0.0001**	-0.0001*
				(0.0001)	(0.0001)	(0.0001)
Dist. to border*(Year of sale-2003)		0.0006***	0.0006***	0.0006***	0.0006***	0.0006***
		(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Dist. to border*Urban area					0.0024	0.0015
					(0.0032)	(0.0032)
Dist. to border*Travel time Dutch city						0.0006***
						(0.0002)
λ	0.0206***	0.0205***	0.0205***	0.0212***	0.0211***	0.0206***
	(0.0076)	(0.0076)	(0.0076)	(0.0076)	(0.0076)	(0.0076)
Obs.	4,657	4,657	4,657	4,657	4,657	4,657

Note: ***, ** and * denote that the estimated regression coefficients are statistically significant at the 1, 5 and/or 10 percent level. In all the regression models presented we control for an extensive list of property- and neighborhood characteristics equivalent to those presented in table C.3 in appendix C.

Table C.6: *Estimation results distance to the border (spatial error model, inverse-distance spatial weights matrix)*

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Dist. to border	-0.0029	-0.0083***	-0.007*	-0.0242***	-0.0263***	-0.0324***
	(0.0018)	(0.0024)	(0.0041)	(0.0076)	(0.0077)	(0.0081)
Dist. to border^2			-0.0001	0.0029**	0.003***	0.0024**
			(0.0002)	(0.0011)	(0.0011)	(0.0012)
Dist. to border^3				-0.0001***	-0.0001***	-0.0001***
				(0.0001)	(0.0001)	(0.0001)
Dist. to border*(Year of sale-2003)		0.0006***	0.0006***	0.0006***	0.0006***	0.0006***
		(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Dist. to border*Urban area					0.0057*	0.005
					(0.0032)	(0.0032)
Dist. to border*Travel time Dutch city						0.0005**
						(0.0002)
ρ	0.0471**	0.048**	0.048**	0.0482**	0.0483**	0.0479**
	(0.0224)	(0.0225)	(0.0225)	(0.0223)	(0.0221)	(0.0223)
Obs.	4,657	4,657	4,657	4,657	4,657	4,657
R-sq.	0.833	0.833	0.833	0.833	0.834	0.834

Note: ***, ** and * denote that the estimated regression coefficients are statistically significant at the 1, 5 and/or 10 percent level. In all the regression models presented we control for an extensive list of property- and neighborhood characteristics equivalent to those presented in table C.3 in appendix C.

Table C.7: *Estimation results distance to the border (spatial lag model, inverse-distance spatial weights matrix)*

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Dist. to border	-0.0029	-0.0082***	-0.007*	-0.0238***	-0.0259***	-0.0321***
	(0.0018)	(0.0024)	(0.0041)	(0.0076)	(0.0076)	(0.0081)
Dist. to border^2			-0.0001	0.0029**	0.0029**	0.0024**
			(0.0002)	(0.0011)	(0.0011)	(0.0012)
Dist. to border^3				-0.0001***	-0.0001***	-0.0001***
				(0.0001)	(0.0001)	(0.0001)
Dist. to border*(Year of sale-2003)		0.0006***	0.0006***	0.0006***	0.0006***	0.0006***
		(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Dist. to border*Urban area					0.0055*	0.0048
					(0.0032)	(0.0032)
Dist. to border*Travel time Dutch city						0.0005**
						(0.0002)
λ	0.0223***	0.0222***	0.0222***	0.0231***	0.0227***	0.0224***
	(0.0076)	(0.0076)	(0.0076)	(0.0076)	(0.0076)	(0.0076)
Obs.	4,657	4,657	4,657	4,657	4,657	4,657

Note: ***, ** and * denote that the estimated regression coefficients are statistically significant at the 1, 5 and/or 10 percent level. In all the regression models presented we control for an extensive list of property- and neighborhood characteristics equivalent to those presented in table C.3 in appendix C.

Table C.8: *Estimation results travel time to the nearest border crossing (ordinary least squares)*

Variable	(1)	(2)	(3)	(4)	(5)
Travel time border	-0.003*	-0.0081***	-0.0146***	-0.0152***	-0.0153***
	(0.0016)	(0.0022)	(0.0034)	(0.0034)	(0.0039)
Travel time border^2			0.0003**	0.0003**	0.0003
			(0.0001)	(0.0001)	(0.0003)
Tr. time border*(Year of sale-2003)		0.0006***	0.0006***	0.0006***	0.0006***
		(0.0002)	(0.0002)	(0.0002)	(0.0002)
Tr. time border*Urban area				0.0015	0.0015
				(0.0025)	(0.0025)
Tr. time border*Tr. time Dutch city					0
					(0.0002)
Obs.	4,657	4,657	4,657	4,657	4,657
R-sq.	0.834	0.835	0.835	0.835	0.835

Note: ***, ** and * denote that the estimated regression coefficients are statistically significant at the 1, 5 and/or 10 percent level. In all the regression models presented we control for an extensive list of property- and neighborhood characteristics equivalent to those presented in table C.3 in appendix C.

Table C.9: *Estimation results travel time to the nearest border crossing (spatial error model, inverse-distance spatial weights matrix)*

Variable	(1)	(2)	(3)	(4)	(5)
Travel time border	-0.0031**	-0.0083***	-0.0146***	-0.0152***	-0.0153***
	(0.0014)	(0.0019)	(0.0031)	(0.0031)	(0.0037)
Travel time border^2			0.0003***	0.0003**	0.0003
			(0.0001)	(0.0001)	(0.0002)
Tr. time border*(Year of sale-2003)		0.0006***	0.0006***	0.0006***	0.0006***
		(0.0002)	(0.0002)	(0.0002)	(0.0002)
Tr. time border*Urban area				0.0017	0.0017
				(0.0023)	(0.0023)
Tr. time border*Tr. time Dutch city					0
					(0.0002)
ρ	0.0467**	0.0491**	0.0483**	0.0484**	0.0484**
	(0.0221)	(0.0224)	(0.0226)	(0.0225)	(0.0225)
Obs.	4,657	4,657	4,657	4,657	4,657
R-sq.	0.834	0.835	0.835	0.835	0.835

Note: ***, ** and * denote that the estimated regression coefficients are statistically significant at the 1, 5 and/or 10 percent level. In all the regression models presented we control for an extensive list of property- and neighborhood characteristics equivalent to those presented in table C.3 in appendix C.

Table C.10: *Estimation results travel time to the nearest border crossing (spatial lag model, 10 nearest neighbors spatial weights matrix)*

Variable	(1)	(2)	(3)	(4)	(5)
Travel time border	-0.0028*	-0.008***	-0.0148***	-0.0146***	-0.0157***
	(0.0014)	(0.0019)	(0.003)	(0.0031)	(0.0037)
Travel time border^2			0.0004***	0.0004***	0.0003
			(0.0001)	(0.0001)	(0.0002)
Tr. time border*(Year of sale-2003)		0.0006***	0.0006***	0.0006***	0.0006***
		(0.0002)	(0.0002)	(0.0002)	(0.0002)
Tr. time border*Urban area				-0.0005	-0.0006
				(0.0023)	(0.0023)
Tr. time border*Tr. time Dutch city					0.0001
					(0.0002)
λ	0.0217***	0.0215***	0.0215***	0.0216***	0.0214***
	(0.0076)	(0.0075)	(0.0075)	(0.0075)	(0.0075)
Obs.	4,657	4,657	4,657	4,657	4,657

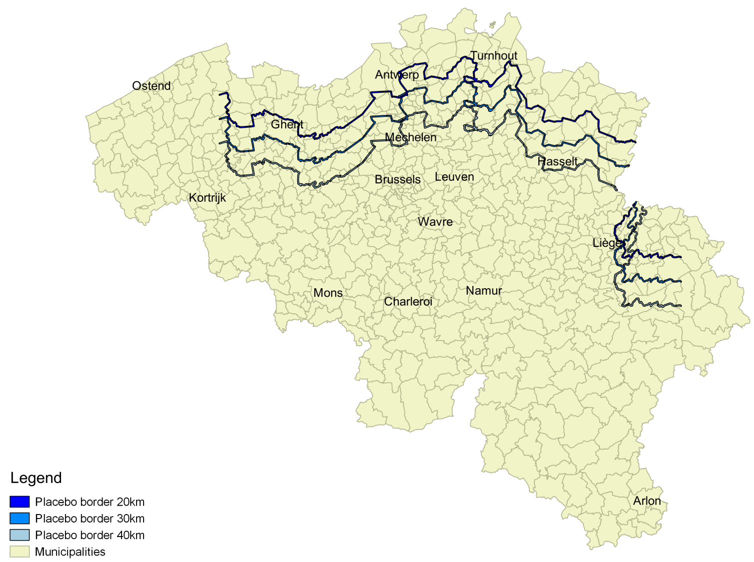
Note: ***, ** and * denote that the estimated regression coefficients are statistically significant at the 1, 5 and/or 10 percent level. In all the regression models presented we control for an extensive list of property- and neighborhood characteristics equivalent to those presented in table C.3 in appendix C.

Table C.11: *Estimation results travel time to the nearest border crossing (spatial lag model, inverse-distance spatial weights matrix)*

Variable	(1)	(2)	(3)	(4)	(5)
Travel time border	-0.003**	-0.0081***	-0.0147***	-0.0152***	-0.0153***
	(0.0014)	(0.0019)	(0.003)	(0.0031)	(0.0037)
Travel time border^2			0.0003***	0.0003***	0.0003
			(0.0001)	(0.0001)	(0.0002)
Tr. time border*(Year of sale-2003)		0.0006***	0.0006***	0.0006***	0.0006***
		(0.0002)	(0.0002)	(0.0002)	(0.0002)
Tr. time border*Urban area				0.0015	0.0015
				(0.0022)	(0.0023)
Tr. time border*Tr. time Dutch city					0
					(0.0002)
λ	0.0234***	0.0232***	0.0232***	0.0231***	0.0231***
	(0.0076)	(0.0076)	(0.0076)	(0.0076)	(0.0076)
Obs.	4,657	4,657	4,657	4,657	4,657

Note: ***, ** and * denote that the estimated regression coefficients are statistically significant at the 1, 5 and/or 10 percent level. In all the regression models presented we control for an extensive list of property- and neighborhood characteristics equivalent to those presented in table C.3 in appendix C.

Figure C.3: *Placebo borders 20, 30 and 40 kilometers*



Source maps: Belgian HISGIS

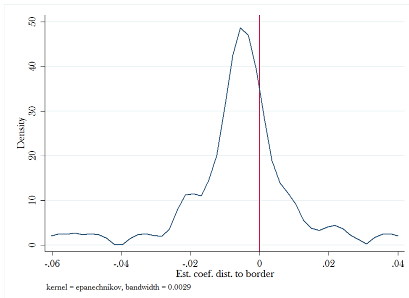
Table C.12: *Estimated coefficients spatial heterogeneity border effect*

Segment length	5km		10km		20km		30km		40km	
Border segment	b_k	s_k	b_k	s_k	b_k	s_k	b_k	s_k	b_k	s_k
0	0.0025	(0.0169)	0.0017	(0.0168)	-0.0037	(0.0076)	-0.0093	(0.0069)	-0.0066	(0.0065)
1			-0.0047	(0.0078)	-0.008	(0.0067)	-0.0055	(0.0078)	-0.0019	(0.0047)
2			-0.0093	(0.0071)	-0.0082	(0.0103)	-0.0025	(0.0041)	-0.0034	(0.0044)
3	-0.0055	(0.0081)	-0.0034	(0.0088)	-0.0003	(0.0048)	-0.0013	(0.0042)	-0.009*	(0.0049)
4	-0.0101	(0.0076)			-0.002	(0.0048)	-0.0086*	(0.0049)	-0.0078	(0.0061)
5	-0.0077	(0.0079)	-0.0067	(0.0105)	-0.0036	(0.0052)	-0.0073	(0.0062)	-0.0238***	(0.0057)
6	-0.0027	(0.009)	-0.0005	(0.0048)	-0.0081*	(0.0049)	-0.0286	(0.0246)	-0.0113**	(0.0055)
7			0.0019	(0.0078)			-0.0233***	(0.0058)	-0.0075**	(0.0033)
8			-0.0034	(0.0054)	-0.0071	(0.0061)	-0.0107*	(0.0057)	-0.0086**	(0.004)
9			-0.0012	(0.005)			-0.0034	(0.0038)	-0.0127**	(0.0058)
10	-0.0075	(0.0107)	-0.0035	(0.0052)	-0.0301	(0.0245)	-0.0094**	(0.0037)	-0.0099	(0.0073)
11	-0.0222	(0.022)			-0.0232***	(0.0058)	-0.015***	(0.0049)		
12	-0.0484*	(0.028)			-0.0105*	(0.0057)	-0.015**	(0.0061)		
13	-0.0008	(0.0049)	-0.0076	(0.0049)	0.0297	(0.0278)	-0.0118	(0.0081)		
14	0.0054	(0.0112)			-0.0036	(0.0038)				
15	0.001	(0.0082)			-0.01***	(0.0037)				
16	-0.0043	(0.0057)			-0.0106**	(0.0042)				
17	-0.003	(0.0055)	-0.0067	(0.0062)	-0.0248	(0.0162)				
18	0.0087	(0.0116)			-0.0216***	(0.0073)				
19	-0.0015	(0.005)			-0.0065	(0.007)				
20	-0.0038	(0.0054)	-0.026	(0.0243)	-0.0043	(0.008)				
21	-0.0029	(0.0056)								
22			-0.0216***	(0.006)						
24			-0.0079	(0.0062)						
26	-0.0091*	(0.0051)								
27	-0.0046	(0.0051)	0.0345	(0.0281)						
28			0.0027	(0.0072)						
29			-0.0048	(0.0039)						
30			-0.0111***	(0.0038)						
31			0.0089	(0.0095)						
32			-0.0104**	(0.0044)						
33			-0.015***	(0.0052)						
34	-0.0102	(0.0085)	-0.032*	(0.0184)						
35	-0.002	(0.007)	-0.0188	(0.0262)						
36			-0.0423	(0.0312)						
37			-0.0222***	(0.0073)						
38			-0.0069	(0.0075)						
39	-0.0226	(0.0236)	-0.0074	(0.0086)						
41			-0.005	(0.0091)						
44	-0.0191***	(0.0061)								
47	-0.0062	(0.0063)								
48	-0.0015	(0.0087)								
55	0.0081	(0.0077)								
58	0.0012	(0.0049)								
59	-0.0095**	(0.0046)								
60	-0.0197***	(0.0044)								
61	-0.0107***	(0.0038)								
62	0.0191*	(0.0109)								
64	-0.002	(0.0098)								
65	-0.0116***	(0.0044)								
67	-0.0127**	(0.0053)								
68	0.0372	(0.0531)								
70	0.0121	(0.0326)								
72	0.0238	(0.0622)								
73	-0.0574	(0.0372)								
74	-0.0333***	(0.0104)								
75	-0.0177**	(0.0076)								
76	-0.0058	(0.0075)								
78	-0.007	(0.0087)								
81	-0.0049	(0.0093)								
82	-0.0044	(0.0167)								
83	0.0055	(0.0178)								
R-sq.	0.837		0.836		0.835		0.835		0.835	
Obs.	4,561		4,561		4,561		4,561		4,561	
ρ	0.1535***	(0.0354)	0.1803***	(0.0344)	0.1871***	(0.0342)	0.193***	(0.0339)	0.1951***	(0.0337)

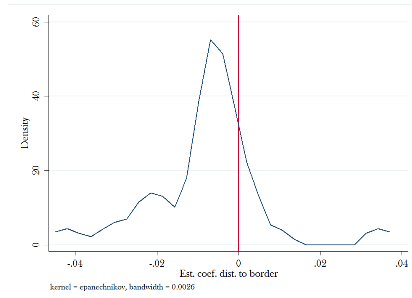
Note: ***, ** and * denote that the estimated regression coefficients are statistically significant at the 1, 5 and/or 10 percent level. In all the regression models presented we control for an extensive list of property- and neighborhood characteristics equivalent to those presented in table C.3 in appendix C.

Figure C.4: *Kernel density plots spatial heterogeneity border effects for various lengths of the border segments*

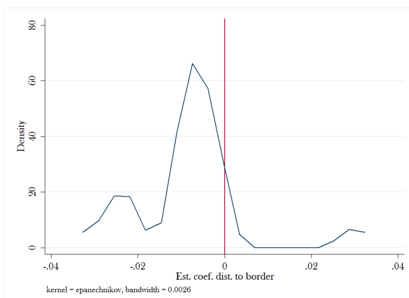
(a) 5km



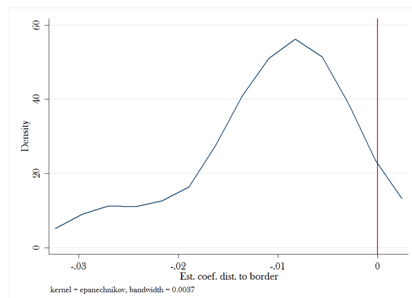
(b) 10km



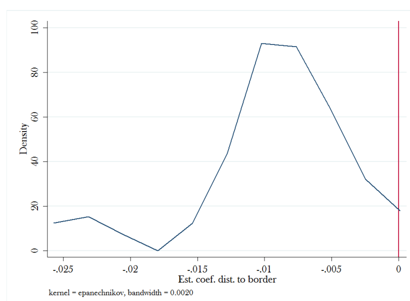
(c) 20km



(d) 30km



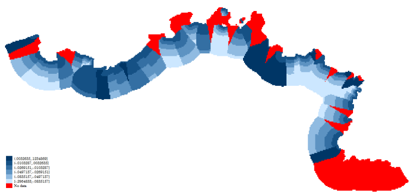
(e) 40km



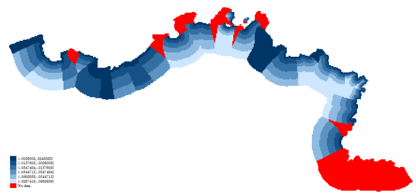
Source maps: Belgian HISGIS

Figure C.5: *Spatial heterogeneity border effects for various lengths of the border segments*

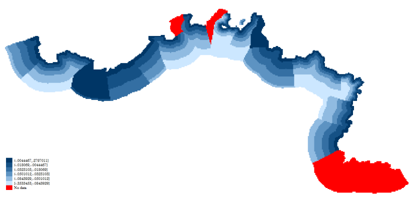
(a) 5km



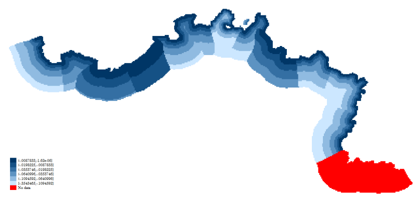
(b) 10km



(c) 20km



(d) 30km



(e) 40km

Source maps: Belgian HISGIS

Chapter III

Housing Prices and Housing Stock Composition: How Do We Value the Homes of our Neighbors?

1 Introduction

The housing commodity is a durable and highly heterogeneous that is fixed in geographical space.[1] Each house has its own unique set of characteristics and amenities that affect its value and, as a result, valuing a house is difficult. A substantial body of historical research has nevertheless attempted to explain value of housing by valuing its individual components. Typically, this has been done using the hedonic pricing method, which assumes that utility is not generated not by the good per se, but by the characteristics that define the good. The view that the housing commodity can be characterized as a "*bundle of attributes*" that provides housing services to its occupants has been generally accepted in the literature and many studies have included the attributes of the property that is up for sale in hedonic regression analyses. At a slightly more aggregate level, however, neighbor-

[1]I would like to thank Jan Mutl, Jan Rouwendal, Frank Verboven Maarten Goos, Erik Buyst and the seminar and conference participants of the EBS Workshop on Real Estate Economics and Finance (Wiesbaden, March 2015), ENHR Housing Economics Workshop (Oslo, March 2015) and (Flemish) Policy Research Centre Housing (Leuven, March 2015) for useful suggestions and comments on earlier drafts of this paper. I would furthermore like to thank ERA Belgium, the General Administration of Patrimonial Documentation (GAPD) and the Flemish Geographical Information Agency (FGIA) for providing the necessary data.

hoods[2] can also be viewed of "*bundles of properties*" where each property has its own specific set of attributes. Although the hedonic pricing literature has clearly shown that different bundles of housing characteristics are valued differently and the neighborhood's amenities matter, there have been virtually no studies that investigated whether different bundles of properties in an otherwise equivalent neighborhood are valued differently. It is however likely that the price of the property that is up for sale is affected by the characteristics of neighboring properties because of reasons related to beauty and aesthetics. Ahlfeldt & Maennig (2010) and Ahlfeldt & Mastro (2012), for example, report positive externalities of historic landmark buildings on surrounding property values, which suggests that not only the characteristics of the property itself matter for its price.

While the aforementioned studies are part of a strand of literature that explicitly examines the spillover effects of certain well-defined buildings on surrounding property values, the current paper focuses on the summary statistics of several well-defined and easily measurable attributes of *all the properties* surrounding the property that is up for sale. More specifically, we investigate whether average values and diversity measures (e.g. standard deviation) of the characteristics of *all* neighboring properties affect the value of the property that is up for sale. We thereby look at the following characteristics: building type, year of construction, size and shape of buildings and plots, and the distance to the road of buildings. We thus try to shed light on the following research question: "*to what extent is the price of a house affected by the characteristics of neighboring homes?*" We therefore combine a large and detailed dataset of approximately 6,100 dwellings that were sold between 2003 and 2015 in the Flemish part of the Brussels Metropolitan Area with administrative and geospatial data concerning the composition of the housing stock at the neighborhood level.[3] [4] [5] The results of a (spatial) hedonic model indicate that the price of an observationally equivalent property is increasing in the average size, year of construction, maximum roof height

[2] Goodman (1978) defines neighborhoods as "*a small urban area within which the residents receive or perceive a common set of socio-economic effects and neighborhood services.*"

[3] The Brussels Metropolitan Area is one of the 18 designated metropolitan/urban areas in Belgium as defined by Luyten & Van Hecke (2007).

[4] Flanders is one of the three federal regions of Belgium, together with Wallonia and the Brussels Capital Region.

[5] Although the transaction data and administrative data are available for the entire Brussels Metropolitan Area, we restrict our attention to the Flemish part of the BMA since the geospatial data are only available for Flanders.

and distance to the road of neighboring properties, but decreases in the average size and shape of the plots of neighboring properties. Our finding that the price of a property is increasing in the average size of neighboring properties provides supporting evidence for the *tax capitalization hypothesis* of Hamilton (1976), who argues that relatively smaller houses benefit from the presence of larger houses. The results presented for the diversity measures reveal that home values are higher in neighborhoods that are characterized by low levels of diversity in (1) the types of buildings, (2) the years of construction, (3) the living areas, (4) the shapes of buildings, and (5) distances to the road. Our findings furthermore indicate that after controlling for an extensive list of property- and neighborhood characteristics, differences in the composition of the neighborhood's housing stock are responsible for price differentials as large as 12%. This finding suggests that not only the characteristics of the property itself matter for its sales price, but also the characteristics of neighboring properties.

While the results provide just one example of how new and ever more present geospatial data can be used in econometric analyses, it is also important to emphasize that the results presented in this paper have important and potentially far-reaching consequences for real estate professionals, policy makers, and urban planners. Our findings, for example, provide backing for the often-heard realtor's advice to buy a small house in a neighborhood that is dominated by larger houses. We, for example, also find that an equivalent property is worth 0.8% more in a neighborhood where the average property is 10 years younger, which suggests that the physical deterioration of neighboring properties has an impact upon the price/value of the property that is up for sale. This consequently implies that renovation subsidies for the exterior of buildings might not only have an impact upon the value and quality of renovated properties, but also have spillover effects on the values of neighboring properties that should be taken into account in their cost-benefit analysis. The finding that households generally prefer to live in neighborhoods that are characterized by low levels of diversity in building types, year of construction, and construction size is of interest for urban planners and real estate developers when thinking about the development of new neighborhoods. At a more general level the current paper shows that data concerning neighboring properties that is becoming ever more available in the form of, for example, geospatial data can be used to improve mass appraisal systems as was also argued by Lindentahl (2016).

The rest of this paper is set out as follows. In section 2, we provide an (non-exhaustive) overview of the literature that is related to the current study. In section 3, we briefly introduce the Brussels Metropolitan Area and describe how the typical Flemish neighborhood is developed and subsequently constructed. In the fourth section the data and estimation methods are presented that are used in the subsequent empirical analyses. In section 5 the results of various empirical analyses are presented and discussed. Finally, section 6 concludes.

2 Literature Review

The current study firstly relates to a long tradition of hedonic price studies in economic research. The hedonic pricing method is based on the observation that goods are valued for their utility bearing characteristics and has become one of the main workhorses of (housing) economists since the seminal contributions of Lancaster (1966) and Rosen (1974). Excellent overviews of the use of the hedonic pricing model in housing economics can be found in Malpezzi (2003) and Sirmans *et al.* (2005). In the current paper we employ hedonic techniques to estimate the marginal willingness-to-pay (WTP) for property- and neighborhood attributes.

The current paper also relates to a strand of literature that has been concerned with the relative consumption of households. Classical economists, such as Adam Smith and John Stuart Mill, already suggested that individuals are at least partly motivated by concerns of relative position. Luttmer (2005), for example, finds evidence that the negative effect of increases in neighbors' earnings on own well-being is most likely caused by interpersonal preferences, that is, people have utility functions that depend on relative consumption in addition to absolute consumption. While the evidence provided by Luttmer (2005) suggests that "*men do not only desire to be rich, but also richer than other men*", in real estate markets, there is no consensus among economists and real estate professionals about the sign and magnitude of this relationship. While the conspicuous consumption theory, which dates back to Veblen (1899), suggests that households have a higher willingness to pay for a large dwelling to signal their supposed affluence to their neighbors, Hamilton (1976) argues that capitalization of the property tax for a given public service bundle penalizes larger houses and benefit smaller houses. A third theory, that can be attributed to Haurin (1988), states that there is a non-monotonic relationship between relative size and housing prices, where more atypical dwellings will sell at a discount. Despite these conflicting hypotheses, there are relatively few empirical studies investigating the relationship between relative size and housing prices. In one of the few attempts, to our knowledge, Turnbull *et al.* (2006) use a simultaneous price-liquidity model and a sample of 2,111 transactions from the Baton Rouge, Louisiana, housing market. They find evidence that smaller houses command a premium over larger houses. In their study, Narwold & Sandy (2010) also find similar evidence. In this paper, we augment the traditional hedonic pricing analysis with the average values of several property characteristics of neighboring properties. The inclusion of both the property's own size and the size of neighboring properties allows us to say something about the value of relative consumption.

Our paper furthermore relates to a strand of literature that has investigated the relationship between diversity and housing values. Although a considerable body of research has primarily focused on the relationship between racial composition and housing prices (e.g. Burnell, 1988; MacPherson & Sirmans, 2001; Collins & Margo, 2003; Cervero & Duncan, 2003), there are a few studies that focus on the role of housing stock diversity. Miller (1978), for example, provides evidence which suggests that houses in more homogeneous neighborhoods tend to have a shorter time on the market than houses in diverse neighborhoods. In a study that is perhaps the most related to ours, Narwold & Sandy (2010) explore the role of housing stock diversity in explaining variation in housing prices. Using hedonic techniques and a sample of approximately 6,500 transactions from the San Diego (California) housing market, they find that dwellings located in more heterogeneous neighborhoods in terms of dwelling size command a premium. Their results, however, also suggest that the price is decreasing with respect to the level of heterogeneity in the age of buildings. Although the current paper is closely related to that of Narwold & Sandy (2010), it also differs in a number of ways. Firstly, the current paper uses data from a different market, namely the Brussels Metropolitan Area. Secondly, and perhaps more importantly, in the current paper we do not only have administrative data at a predefined level at our disposal, but also use geospatial data which allows us to aggregate flexibly and to construct additional variables, such as the shape of buildings and plots. The current paper therefore also relates to a recent study by Lindentahl (2016), who uses a 3D-model for the city of Rotterdam to quantify shape (dis)similarity. His results suggest that shape similarity can be used to improve mass appraisal systems.

This study additionally relates to a strand of literature that has investigated the spillover effects of dwellings designed by famous architects or (historic) landmark buildings/neighborhoods on surround property values. Ahlfeldt & Mastro (2012), for example, show that homes within 50 meters to a residential building by Frank Lloyd Wright in Oak Park, Illinois, enjoy a price premium of 8.5%. The authors, however, argue that it may also be partially attributable to the prominence of the architect. More generally, positive externalities of (historic) landmark buildings and designated areas have been documented for Texan cities (Leichenko *et al.*, 2001; Coulson & Leichenko, 2004), Baton Rouge (Zahirovic-Herbert & Chatterjee, 2012), Memphis (Coulson & Lahr, 2005), Berlin (Ahlfeldt & Maennig, 2010), the Netherlands (Lazrak *et al.*, 2014), England (Ahlfeldt & Holman, 2015), and the United States in general (Listokin *et al.*, 1998). While all these studies look at the spillover effects of well-designated buildings/areas, the current study looks more generally at the effect of the composition of the neighborhood's housing stock on housing prices.

Our paper finally relates to a strand of literature which investigates the effects of architecture on market prices of residential and non-residential buildings. Moorhouse & Smith (1994), for example, investigate the market for residential architecture for row houses in Boston's South End. Their results show that "*residential architecture matters in the marketplace*" and that "*specific architectural features are more highly valued when they differentiate one row house from its immediate neighbor.*" Fuerst *et al.* (2011) investigate whether commercial offices designed by 'signature architects' in the United States achieve rental premiums compared with commercial offices designed by non-signature architects. The results from their hedonic regression analysis suggest that, compared with buildings in the same submarket, office buildings designed by signature architects have rents that are 5%-7% higher, and sell for prices 17% higher. Although we do not explicitly take into the architectural features of the properties that are up for sale in the current paper, we do include shape variables for the construction and the plot is sits on in the empirical analysis.[6]

[6] The construction of the shape measures is described in appendix A.3.

The current paper finally closely relates to the literature on spatial econometrics. Including the characteristics of neighboring properties in a hedonic pricing model is highly similar to the SLX model (e.g. Halleck Vega & Elhorts, 2015), where spatially lagged values of independent variables are included as (additional) regressors in the regression analysis. The approach in the current study, however, deviates somewhat from this strand of literature as the spatially lagged values are constructed using data for the whole housing stock instead of only using data from observations from the sample. We furthermore use Spatial Lag (SAR) and Spatial Error Models (SEM) to estimate the hedonic coefficients.

3 The Brussels Metropolitan Area and institutional context

3.1 The Brussels Metropolitan Area

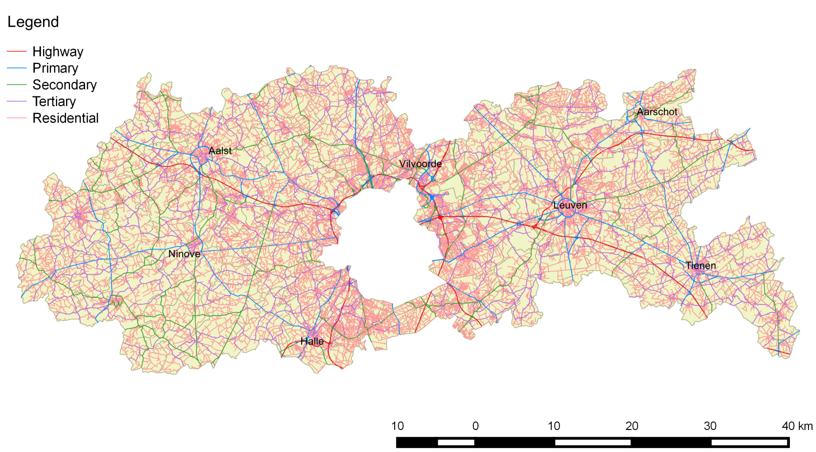
In this paper we investigate the relationship between neighborhood housing stock composition and housing prices using data from the *Flemish* part of the larger Brussels metropolitan area (BMA).[7] [8] Due to data limitations, which are discussed in section 4, we only consider the Flemish part of the Brussels metropolitan. We thus exclude transaction from the Brussels Capital Region (BCR), which is located in the heart of our study region, and the Walloon part of the BMA. The aforementioned Brussels Capital Region is one of the three federated regions of Belgium alongside Wallonia and Flanders and is, besides the capital of Belgium, also the *de facto* capital of the European Union since it hosts a number of principal EU institutions, such as the European Commission and the Council of the European Union. The secretariat of the Benelux and the headquarters of the North Atlantic Treaty Organization (NATO) are also located in Brussels, as well as many other international organizations such as the World Customs Organization,

[7] We use the definition set out by Luyten & Van Hecke (2007), who divide Belgium into 5 large metropolitan areas (Brussels, Antwerp, Ghent, Liège and Charleroi) and 13 smaller urban areas (Mons, Leuven, Bruges, Namur, Kortrijk, Mechelen, Hasselt, Verviers, Ostend, Tournai, Genk, Sint-Niklaas and Turnhout) based on socio-economic data.

[8] We refer to the larger BMA since we also retain transaction in the smaller Leuven metropolitan area. Although Leuven is a separate urban area according to Luyten & Van Hecke (2007), it is encompassed by the Brussels metropolitan area in the east, south and west.

EUROCONTROL and a large number of international corporations. It is no wonder then the BCR is Belgian's largest center of employment[9] and also responsible for a large influx (outflow) of employees every morning (evening) from our study region and the rest of Belgium. Data from the Census 2011, a large-scale questionnaire carried out by Statistics Belgium, suggests that of the 575,000 people that are employed in our study region, more than 150,000 ($\approx$ 27%) are employed in the BCR, which suggests that the travel time to Brussels will have an important effect on housing prices and thus has to be included in the hedonic pricing analysis.[10] Other centers of employment in the Flemish part of the BMA are Leuven ($\approx 72,000$), Zaventem ($\approx 42,000$) and Aalst ($\approx 34,000$). In figure 1 a geographical overview of the study region is presented.

Figure 1: *Geographical overview study region*



Source maps: Belgian HISGIS

3.2 Institutional context

As it is the goal of this paper to investigate the relationship between neighborhood housing stock composition and housing prices, we cannot continue without a brief description of how neighborhoods have come about in Flan-

[9]Approximately 675,000 people are employed in the BCR. About 388,000 of them commute (on a daily basis) from municipalities outside the BCR. *Source:* Census 2011, Statistics Belgium.

[10]In the hedonic pricing analyses presented in section 5 we explicitly control for the travel time to Brussels. The travel time was calculated using the *osrmtime* algorithm in Stata 14 (Huber & Rust, 2016), which calculates travel time and distance using OpenStreetMap (OSM) data and an Open Source Routing Machine (OSRM).

ders in the first place. As, for example, argued in Tennekes *et al.* (2015), it is intuitive that the morphology of residential development is influenced to a great extent by institutional arrangements. According to Ball (2003) and Verhage (2002), the development process contains at least three stages that are interrelated, namely 1) land assembly and development, 2) house construction, and 3) house sales/end use. Differences in the institutional arrangements between countries in these three phases can give rise to discontinuities in residential morphology along national borders. Tennekes *et al.* (2015), for example, note that "*it is still relatively easy to see the lines of the border when crossing from the Netherlands to Belgium [...] due to differences in dwelling types and the shape of urban development.*" In a comparative study Tennekes *et al.* (2015) show that the housing process in Flanders is quite different from that in the Netherlands and North-Rhine Westphalia (Germany). In Flanders, an excess of land that was zoned for residential purposes (De Decker, 2011; Loris, 2011), combined with a highly fragmented possession of the available land plots (Loris, 2008), has led to a situation where generally the landowner took the initiative for development, either individually or by selling it to a property developer (*verkavelaar*). Consequently, the construction of new dwellings and apartments was, and still is, frequently organized on a small-scale basis in Flanders/Belgium. A quick glance at *www.immoweb.be*[11], the largest listing platform in Belgium, reveals that the 20 largest developers of (residential) real estate in Belgium are listing 3,569 properties spread over 399 different projects, which suggests that the average development consists of 8.94 properties. Data retrieved from the website of *Statistics Belgium* furthermore suggests that the combined market share of the 20 largest construction companies active on *www.immoweb.be* is fairly limited (6.6%) as there were granted building permits for more than 54,000 residences in Belgium in 2014. In contrast, the largest housing development in the Netherlands, *Leidsche Rijn* near the city of Utrecht, consists of 30,000 homes that were simultaneously developed and built in a single development. This image of small-scale construction of new housing developments in Flanders is also confirmed by data from the Flemish Housing Survey 2013, a large-scale questionnaire where 10,000 randomly selected Flemish households were interviewed about their housing expenditures, methods of financing, their housing needs, and the housing career.

[11]Status at the 16th of August 2015.

More than half[12] of all the respondents who reported that they bought or constructed a new residence since 1990, reported that they individually constructed their own house with the help of one or multiple architects and/or contractors.

The image of Flanders (or more generally Belgium), with a lack of spatial planning in the first decades after the second World War, highly fragmented possession of open residential lots, and a large share of owner-occupiers who designed and built their own dwellings, suggests that neighborhoods are quite likely to be heterogeneous in terms of their housing stock. This is in contrast to, for example, the Netherlands where new housing developments are frequently organized on a large scale and dwellings in the same neighborhood are frequently much alike. This raises the question: "*What do households prefer*?" In the next sections we try to shed some light on the answer to this question.

4 Data & methodology

4.1 Data

In this paper, we combine information from four main sources. The main dataset used in the subsequent empirical analyses is a sample of 6,450 dwellings that were sold through one of 50 real estate agencies of a large franchise system in the Flemish part of the Brussels and Leuven metropolitan areas in the period 2005-2014.[13] For every property in the dataset, we know its transaction price (in €), date of sale, construction type (terraced vs. (semi-)detached), interior space (in m^2), plot size (in m^2), # of bedrooms, # of bathrooms, and year of construction. The dataset furthermore comprises a wide range of dummy variables that indicate whether certain

[12] Of the total sample of 10,013 respondents, 7,561 (75.51%) are owner-occupiers. Of these 7,561 owners-occupiers, 3,119 (41.2%) reported that they had either bought a newly constructed house or constructed a new house themselves. Of these 3,119 respondents 2,180 (69.9%) reported that they constructed their own house with or without the help of one or multiple architects and/or contractors.

[13] The dataset used for the current analyses is part of a larger sample of approximately 55,000 transactions of dwellings, apartments, offices, garages and residential land that were sold through one of 110 real estate agencies of the franchise system in Belgium in the period 2003-2015. For the sake of homogeneity and easier access for real estate agents to other agents' listings, the franchise system introduced and manages a centrally stored where all listings are registered, together with their actual status. The real estate agents/agencies are obliged and trained to submit the characteristics of properties of clients in a highly structured manner, which results in a very clean database.

features (e.g. central heating, dishwasher) are present or not and whether certain criteria are met (e.g. condition = ready to move in).[14] More importantly for the current analysis, the dataset also contains the address (street, house number, zip code) and corresponding geographical coordinates (e.g. 50.87 °,4.70 °), which allows us to merge the data with the neighborhood variables and housing stock diversity measures that are discussed in the next paragraphs. In table 1 some descriptive statistics are provided.

Table 1: *Descriptive statistics transaction data*

Variable	Avg.	St. Dev.	P5	P95
Sales price (in €)	260,272.33	105,123.11	121,468	450,000
Semi-detached	0.3	0.46	0	1
Detached	0.41	0.49	0	1
Living area (in sq. m.)	198.38	72.68	110	338
Plot size (in sq. m.)	788.49	1,057.04	112	2,362
# Bedrooms	3.17	0.96	2	5
# Garages	0.91	0.71	0	2
Year of constr.	1965.46	21.52	1930	2005
Year of sale	2009.64	3.59	2003	2015
Travel time Brussels (in min.)	21.58	8.44	11.08	36.65

Note: the descriptive statistics were calculated using the sample of 6,450 transactions.

The descriptive statistics reveal that the average property in our sample is sold for €260,272, ranging from €121,468 at the 5^{th} percentile to €450,000 at the 95^{th} percentile. Approximately 71% of all dwellings are either semi-detached (30%) or detached (40%). The average travel time to Brussels by car (in min.) is equal to 21.5 minutes, but ranges from 11 minutes at the 5^{th} percentile to 37 minutes at the 95^{th} percentile. While the table presented above shows that there is considerable variation in the different variables across properties, Moran's *I* statistics (Moran, 1950) for the different variables presented in table A.2 in appendix A.1 show that the values of the variables are also (strongly) correlated across space, which motivates the use of spatial econometric estimation techniques in subsequent empirical analyses.

The transaction data discussed in the previous paragraph is complemented

[14] A list of all the variables at our disposal is presented in table A.1 in appendix A.

with a single cross-section (January 1, 2012) of administrative data concerning the inventory of the Belgian housing stock that was provided by the General Administration of Patrimonial Documentation (GAPD) for every statistical sector[15] in Belgium. In our study area, there are 2,240 statistical sectors with an average territory of approximately one square kilometer and 600 inhabitants. This suggests that it is a suitable neighborhood definition. More specifically, for every statistical sector, we observe the number of buildings/residences that belong to one of the 14 different[16] categories. For every category of buildings/residences, additionally, information is provided concerning some of their structural characteristics, such as interior space, parcel size, built-up area, # of floors, and the year of construction. The data is organised in such a manner that for every local living area j and category of buildings/residences we know the number of buildings/residences that belong to each of the K classes, which are bounded by pre-defined threshold values. An overview of these threshold values is provided in table A.3 in appendix A.2. Using this dataset, we can calculate (approximate) weighted averages[17] for the variables living area, plot size and built-on surface (in m^2), and the year of construction for every local living area j. These weighted averages can be used to capture the "average effect". The dataset also allows us to calculate several diversity measures for the different characteristics l, such as the maximum share, an (approximate) standard deviation and Simpson's (1949) diversity index D. The latter is defined as:[18]

$$D = 1 - \Sigma_{k=1}^{K}(p_k)^2 \qquad \text{(III.1)}$$

where p_k is the proportion of the population within a certain category.[19] While Simpson's (1949) diversity index D has frequently been applied in life

[15] The statistical sectors in Belgium are the smallest administrative level at which Statistics Belgium collects (socio-economic) data. There are 19,781 statistical sectors in Belgium, of which 9,182 are located in Flanders.

[16] (1) Apartments, (2) garages/storage, (3) *buildings* (apartment building owned by a single owner), (4) homesteads/farm houses, (5) trading houses, (6) attached/terraced houses, (7) semi-detached houses, (8) detached houses, (9) offices, (10) castles, (11) industry buildings, (12) social profit buildings, (13) vacation homes, and finally (14) villa's.

[17] The weighted average of characteristic l is equal to $\Sigma_{k=1}^{K} p_k * (\frac{v_k^u - v_k^l}{2})$, where p_k denotes the proportion of the population in class k, and v_k^l and v_k^u denote the lower and upper threshold value for that respective class.

[18] Bivariate correlation statistics between the different diversity measures are provided in table A.4 in appendix A.2.

[19] Note that Simpson's diversity index D is inversely related to the Herfindahl-Hirschman Index that is frequently used as an indicator of the degree of competition among firms.

science applications, this is one of the first applications to housing prices.[20] Descriptive statistics for Simpson's D and other diversity measures are provided in table 2.

Table 2: *Descriptive statistics housing stock diversity*

Variable	Measure	Avg.	St. Dev.	P5	P95
Building types	D	0.613	0.156	0.29	0.776
	Max	0.526	0.162	0.319	0.838
Living area	COV	0.415	0.063	0.325	0.532
	D	0.548	0.077	0.399	0.647
	Max	0.559	0.094	0.446	0.746
Plot size	COV	0.554	0.176	0.292	0.84
	D	0.63	0.143	0.336	0.78
	Max	0.493	0.163	0.299	0.805
Year of constr.	COV	0.015	0.004	0.008	0.021
	D	0.805	0.088	0.643	0.876
	Max	0.299	0.12	0.166	0.539

Note: all statistics were calculated for the 2,240 statistical sectors in our study region.

The descriptive statistics presented above, for example, suggest that the average neighborhood in our study region is composed out of multiple building types, since the average value for the diversity index D is not equal to 0. Note, however, that its standard deviation is also not equal to zero, which suggests that there is across-neighborhood variation that can be exploited to capture the "diversity effect". Although the administrative data cover various characteristics, there are also a few drawbacks. Firstly, the administrative data only contain information on the number of properties of a certain type in a certain class for the characteristics, and these classes are sometimes broadly defined (e.g. plot size is larger than $664m^2$ but smaller than $1304m^2$). Secondly, and perhaps more importantly, the administrative data are gathered at a pre-defined level of aggregation (statistical sectors) and thus do not allow for flexible aggregation. Therefore, we now turn to the geospatial data.

The individual transaction data and administrative data is then further complemented with variables that were created from geospatial data provided by

[20] The first application, to our knowledge, is Narwold & Sandy (2010).

the Flemish Geographical Information Agency (FGIA).[21] From the geospatial data we were not only able to construct several interesting variables for the properties in our transaction data, but were also able to calculate the average values and standard deviations for these variables for neighboring properties. Moreover, we were able to calculate these variables for several levels of aggregation. More specifically, we calculated the *surface* (in m^2), *perimeter* (in m) and *shape*[22] of the plot and built up area for both the own property and neighboring properties. We were furthermore able to retrieve the roof height (in m) and minimum distance to (in m)- and type of road for both the own property and neighboring properties. For the neighboring properties we calculated both the average values and the standard deviations for these characteristics for four different levels of aggregation, namely (1) statistical sectors, (2) properties within a 100 meter radius, (3) street, and (4) 10 nearest neighbors. The descriptive statistics are presented in table 3.

Table 3: *Descriptive statistics neighboring properties constructed from geospatial data (100 meter)*

Type	Char.	Statistic	Obs.	Avg.	St. Dev.	P5	P95
Building	Dist. to road	Avg.	6,404	5.298	5.312	0.347	11.56
		St. dev.	6,389	4.634	4.231	0.879	11.98
	Height	Avg.	6,404	8.784	1.406	6.779	11.09
		St. dev.	6,389	1.738	0.653	0.864	2.794
	Number	Obs.	6,404	39.36	28.48	8	97
	Perimeter	Avg.	6,404	52.85	10.36	40.01	68.92
		St. dev.	6,389	17.79	13.35	6.339	41.21
	Shape	Avg.	6,404	1.152	0.065	1.061	1.264
		St. dev.	6,389	0.126	0.053	0.047	0.215
	Surface	Avg.	6,404	142.7	66.77	84.66	221.5
		St. dev.	6,389	95.85	133.1	24.78	282.6
Plot	Number	Obs.	6,405	47.58	29.67	14	107
	Perimeter	Avg.	6,405	123.7	44.55	72.87	194.9
		St. dev.	6,404	62.59	40.64	23	120.1
	Shape	Avg.	6,405	1.276	0.115	1.106	1.476
		St. dev.	6,404	0.249	0.122	0.1	0.46
	Surface	Avg.	6,405	810.8	1,059	222.9	1,834
		St. dev.	6,404	825.2	1,105	159.3	2,157

[21] An elaborate description of the shapefiles used to construct the different variables used in the regression analyses is provided in appendix A.3

[22] The shape variable was constructed using the following formula: *Perimeter* (in m)$/(4 * \sqrt{\text{Surface (in } m^2)})$

The descriptive statistics presented in table 3 suggest that there is not only considerable variation in both the average values and the standard deviations of characteristics of neighboring properties, which implies that we can identify both the average- and the diversity effect. Notice on the one hand for example that the average value of average distance to the road of neighboring properties for all properties in our transaction is equal to 5.3 meters. The standard deviation of the average distance to the road of neighboring properties is also equal to 5.3, which suggests that there is considerable variation in the average distance to the road that can be exploited to identify the "average effect". The average of the standard deviation of the distance to the road, on the other hand, which is equal to 4.6, shows that there is considerable heterogeneity in the distance to the road of neighboring properties for most properties in our dataset. The standard deviation, which is equal to 4.2, suggests that there is variation in the variation of neighboring properties' distance to the road that can be used to identify the "diversity effect".

While the geospatial data provides detailed information on some of the characteristics of *every* neighboring property and allows us to flexibly aggregate, there are less variables available. It, for example, does not contain information on the type of building, interior space, and year of construction. In the subsequent empirical analyses, we will therefore combine data from both sources to overcome the drawbacks related to each one of them.

4.2 Methodology

Given the nature of our dataset, i.e. individual transaction prices and property characteristics, and consistent with the literature, we employ the Hedonic Pricing Method (HPM) for which the theoretical foundations were established in the seminal works of Lancaster (1966) and Rosen (1974). The basic premise of the HPM is that utility is not generated by the good per se, but by the characteristics that define the good. Malpezzi (2003), in his excellent literature review[23], illustrates the direct applicability to housing perfectly by stating that "*I'm happy to be home, not so much to be in anything called a 'house', so much as to be in a warm dry place, with a quiet space for a comfortable chair, a functioning toilet or a hot bath should I require them, and some other rooms in the house to store stacks of paper*

[23] Another excellent literature review on the use of hedonic pricing models in housing economics is Sirmans *et al.* (2005).

or noisy children." Traditionally, the HPM is estimated using regression analysis where the estimated coefficients represent the implicit prices of the attributes. In the current paper we allow for a more flexible form and estimate variants of the following regression equation:

$$\ln(p_{ijt}) = \underbrace{\beta X_{ijt} + \alpha N_{jt} + \delta I_t}_{\text{Classical hedonic pricing model}} + \overbrace{\gamma \bar{X}_j + \kappa\sigma(X_j)}^{\text{Neighborhood housing stock}} + \underbrace{\lambda \Sigma_{n=1}^{N} w_{in} \ln(p_n)}_{\text{Spatial lag component}} + u_{ijt}$$

where

$$u_{ijt} = \underbrace{\rho \Sigma_{n=1}^{N} w_{in} u_n + \epsilon_{ijt}}_{\text{Spatial error component}} \tag{III.2}$$

where $\ln(p_{ijt})$ denotes the logarithmically transformed sales price of property i located in statistical sector j that is sold at time t. X_{ijt} and N_{jt}, respectively, denote vectors of structural and neighborhood characteristics for each property and I_t is a vector of indicator variables that capture time-specific effects. β, α and δ denote their respective vectors of shadow prices. We then augment the classical hedonic pricing model with average values of the characteristics of neighboring properties and also include measures that capture the heterogeneity in these characteristics for neighboring properties. More specifically, we add (natural logarithms of) averages and heterogeneity measures for the following characteristics: types of buildings, year of construction, interior space, plot size, building shape, plot shape, maximum roof height, and the distance to the nearest road.

While the regression analysis presented so far could be carried out using Ordinary Least Squares (OLS), the regression coefficients could be biased due to spatial autocorrelation present in the dependent and independent variables. We therefore potentially allow for spatial autocorrelation in the dependent variable, $\ln(p_{ijt})$, and spatial autocorrelation in the residuals, u_{ijt}. The spatial structure of the model is captured in a spatial weights matrix, where the elements w_{in} represent the spatial dependency between observation i and observation n and are equal to zero whenever $i = n$.[24] More specifically, we construct a row-normalized[25] *10 nearest neighbors* and a row-normalized *inverse-distance* spatial weights matrix. All models are

[24]An observation cannot be dependent upon itself.
[25]This is conventional in the spatial econometrics literature.

estimated using the *Generalized Spatial 2 Stage Least Squares* (GS2SLS) estimator in Stata 14 that was proposed by Kelejian & Prucha (1998).[26]

5 Results

5.1 Baseline results

In table B.1 in appendix B the results of a hedonic model *without* size and diversity measures for neighboring properties are presented. The coefficients of the house- and neighborhood-specific attributes, such as structure size, plot size, age, bedrooms, bathrooms, travel time to Brussels are familiar and in general have the expected signs and are (highly) significant. The coefficients of determination (R^2) are equal to 0.808 for a model without municipality fixed effects and 0.848 for a model with municipality dummies. 80.8% and 84.8% of all variation in housing prices can thus be explained by a traditional hedonic model. These values serve as a reference point to assess the incremental explanatory power of models with average and diversity measures which are presented below.

In table 4 the results of various models with average values of and/or heterogeneity measures for neighboring properties are presented. All the models presented allow for spatially autocorrelated error terms (SEM) and a *10 nearest neighbors* matrix is used to construct the spatial weights matrix W.[27] It is furthermore necessary to note that we used a 100 meter radius to construct the measures from the geospatial data.[28] In columns (1)-(3) we gradually augment the classical HPM with averages and/or diversity measures of neighboring properties, which culminates in the full model presented in column (4). In column (5), we additionally add dummy variables for the 72 municipalities in our study region.

[26]The models could also be estimated using a Maximum Likelihood estimator, but the GS2SLS estimator additionally allows for heteroskedasticity and is superior in terms of speed. Excellent discussions of the generalized method of moments and instrumental variables estimation approach underlying the GS2SLS-estimator can be found in Arraiz *et al.* (2010) and Drukker *et al.* (2013).

[27]In appendix B robustness checks with respect to the model specification (SAR-model) and the spatial weights matrix (inverse-distance) are presented.

[28]In section 5.2 we carry out robustness tests with respect to the level of aggregation.

Table 4: *Baseline results: average values and diversity measures*

Category	Variable	(1)	(2)	(3)	(4)	(5)
Ln(avg.)	Living area	0.2837***		0.3328***	0.2859***	0.2849***
		(0.0309)		(0.0327)	(0.0335)	(0.0307)
	Plot size	-0.0683***		-0.0371*	-0.0195	-0.0312
		(0.0196)		(0.021)	(0.0212)	(0.0197)
Avg.:	Year of construction	0		0.0007*	0.0011***	0.0008***
		(0.0003)		(0.0003)	(0.0003)	(0.0003)
	Shape buildings				0.0327	0.0926
					(0.0779)	(0.0757)
	Shape plots				-0.1493***	-0.1326***
					(0.0383)	(0.0364)
	Height				0.0148***	0.0168***
					(0.0028)	(0.0027)
	Dist. to road				0.0059***	0.0061***
					(0.0015)	(0.0014)
D:	Buildings		-0.0536**	-0.0755***	-0.0719***	-0.0366
			(0.0254)	(0.0257)	(0.0258)	(0.0232)
	Year of construction		-0.0303	-0.0986**	-0.0936**	-0.0587
			(0.0392)	(0.0403)	(0.0402)	(0.0367)
	Living area		0.0699	-0.1152**	-0.1137**	-0.1776***
			(0.045)	(0.0511)	(0.0508)	(0.0469)
	Plot size		-0.0309	0.061	0.0741*	0.0952***
			(0.0342)	(0.038)	(0.0378)	(0.0337)
St. dev.:	Shape buildings				-0.1843**	-0.2042***
					(0.0759)	(0.0732)
	Shape plots				0.056**	0.0403
					(0.0272)	(0.026)
	Height				-0.0025	0
					(0.004)	(0.0038)
	Dist. to road				-0.0019*	-0.0027***
					(0.001)	(0.001)
Diagnostics:	ρ	0.5454***	0.554***	0.5388***	0.5299***	0.2134***
		(0.0187)	(0.0185)	(0.019)	(0.0193)	(0.032)
	Obs.	4,276	4,276	4,276	4,276	4,276
	R-sq.	0.82	0.816	0.823	0.826	0.86
	Mun. FE:	No	No	No	No	Yes

Note: Standard errors are reported in parentheses. *, ** and *** indicate $p < 0.10$, $p < 0.05$ and $p < 0.01$, respectively. Besides the reported coefficients, all the regression analyses presented furthermore include an extensive list of property- and neighborhood characteristics that is equivalent to those presented in table B.1 in appendix B.

Average values of characteristics of neighboring properties

The results presented in table 4 suggest that the price of an observationally equivalent property is increasing in the average interior space, year of construction, maximum roof height and distance to the road of neighboring properties, but is decreasing in the average size and shape of the plots of neighboring properties.

The estimates presented in column (5), for example, reveal that a 1 percent increase in the interior space of neighboring properties is associated with a 0.285 percent increase of the price of the property that is up for sale. This is consistent with the *tax capitalization hypothesis* of Hamilton (1976), who shows that local property tax differentials can lead to price differentials between large and small houses in heterogeneous communities, where small houses have higher unit prices than larger houses.[29] The critical reader might argue that Hamilton's capitalization hypothesis only applies when properties enjoy the same bundle of public services. In the regression analyses, however, we control for various other neighborhood characteristics, such as income, population density and distance to the city hall. In column (5) we also augment the model with dummy variables for the different municipalities. Since the municipalities are the main provider of public services, we can safely rule out alternative explanations. This result is furthermore consistent with evidence provided by Turnbull *et al.* (2006) and the often-heard realtor's advice to buy a small house in a neighborhood dominated by larger houses. The estimated coefficients similarly show that an equivalent property is worth 0.8 percent in a neighborhood where the average property is 10 years newer. Although the model does not contain information about the state of neighboring dwellings, this result likely suggests that the physical deterioration of neighboring properties has an impact upon the price of the property that is up for sale. A potential policy implication of this finding is that a cost-benefit analysis of renovation subsidies should not only include the positive effects upon the quality of renovated properties, but should also take into account the spillover effects on neighboring properties. The estimates presented in column (5) of table 4 furthermore suggest that an increase in the average maximum roof height with one meter is associated with a 1.7 percent price increase and an increase in the average distance to the road of neighboring properties with one meter is leads to a 0.6 percent price increase. A plausible explanation for the first finding is that the maximum roof height is higher in more densely populated areas (e.g. apartment buildings, town houses) which are typically associated with higher housing prices. Although we explicitly control for the population density of the statistical sector in our model, it might be that this effect is not yet fully accounted for. Moreover, the population density variable is only available for the statistical sectors, which have an average size of one square kilometer. The average maximum roof height, however, is calculated only

[29] It should be noted, however, that the result of Hamilton (1976) is derived using a number of assumptions. Like in the seminal Tiebout (1956) model, Hamilton (1976) assumes that households are perfectly mobile. Hamilton (1976) furthermore states that market interference (e.g. zoning) is required to ensure efficiency.

for all properties within a radius of 100 meters, which implies a surface of approximately 0.031 square kilometers. A likely explanation for the second effect is that households benefit from the presence of open space in front of their house and neighboring properties. While the critical observer might note that properties with a higher value of distance to the road are typically larger, we argue that these potential "size effects" are already accounted for due to the inclusion of (average) the average plot- and building size in our model.

The results presented in table 4, however, also suggest that there are negative effects related to the average size and shape of plots of neighboring properties, albeit that the estimated coefficient of average size (in m^2) of the plots of neighboring properties is no longer statistically significant in columns (4) and (5). The estimates presented in column (5) reveal that an increase in the average shape of neighboring plots with 0.45 is associated with a 6 percent price decrease.[30] People thus prefer and are willing to pay to live in neighborhoods with compact plots. Furthermore note that this finding is not the result of the property' own plot shape, since this variable is already included in our list of control variables.[31]

Diversity measures

Besides the average values for the characteristics of neighboring properties, the regression analyses presented in table 4 also include several diversity measures for the same set of characteristics. On the one hand, the estimated coefficients generally reveal that home values are higher in neighborhoods that are characterized by low levels of diversity in (1) types of buildings, (2) year of construction, (3) interior space, (4) shape, and (5) distance to the road of buildings. Heterogeneity in plot size, on the other hand, positively affects housing prices.

The estimated coefficients for Simpson's D for the types of buildings in columns (2)-(4) suggest that an increase in the heterogeneity of building types has a negative effect on home values. An increase of Simpson's D with 0.1 in the model presented in column (4), for example, leads to a 0.72

[30] This roughly corresponds with comparing perfectly rectangular plots of $625m^2$ with equally sized plots with a width of 10 meters and depth of 62.5 meters.

[31] The estimates for the building/plot shape of the property and the type of road it is located along are presented in table B.2 in appendix B.

percent price decrease. Notice that the estimated coefficient is no longer statistically significant after controlling for municipality fixed effects in column (5). Nonetheless, the sign of the coefficient remains the same and the associated t-value (-1.58) of the estimate is close to being statistical significant at the 10 percent level after the inclusion of municipality fixed effects. A similar story applies for diversity in the years of construction of buildings, where the estimated coefficient is negative and statistically significant in columns (2)-(4) but loses its statistical significance in column (5). The estimated coefficient in column (5) suggests that an increase in Simpson's D with 0.1 is associated with a 0.59 percent price decrease, but we cannot reject the hypothesis that the estimated coefficient is equal to zero. Potential explanations for this loss of statistical significance after the inclusion of municipality fixed effects are (1) that the diversity measures are correlated with (unobserved) municipality effects, and (2) a loss of statistical power due to the inclusion of 71 additional dummy variables.[32] All in all, these results suggest that households have a willingness to pay to live in neighborhoods that are characterized by low levels of diversity in building types and the age of buildings.

While the results presented in the previous paragraph provide a mixed picture, the reported results for the diversity measures of living area, plot size, building shape, and distance to the road are fairly stable across the different regression specifications and thus remain statistically significant even after the inclusion of municipality fixed effects. One the one hand, an increase in Simpson's D with 0.1 for the living area of properties in column (5) is, for example, leads *ceteris paribus* to a decrease of home values with 1.8 percent. An increase in Simpson's D with 0.1 for the plot size, on the other hand, leads to an increase of property values with 0.95 percent. An increase in the standard deviation of the shapes of neighboring buildings with 0.1 results in a price decrease of approximately 2.04 percent.[33] Note, by comparing the estimated coefficient for the average shape of buildings with that of the standard deviation of building shapes, that households do not care about the average shape of buildings, but (negatively) value heterogeneity in the

[32] Essentially, the identification now stems from *within* municipality variation in the diversity measures. It should therefore be noted that our transactions are located in 1,139 statistical sectors of 71 municipalities, which implies that on average there are only 16 statistical sectors in a municipality.

[33] In a neighborhood where the average plot is $600m^2$ with a width of 15 meters and depth of 40 meters, the typical deviation from the average is a plot that has a width of approximately 12.5 meters and depth of approximately 47.5 meters.

shapes of neighboring buildings. Households thus dislike dissimilarity in the shapes of buildings in the neighborhood. The estimates finally reveal that this result also applies for the diversity in the distances to the road of buildings. The estimated coefficient in column (5) indicates that an increase of the standard deviation with one meter leads to a 0.27 percent price decrease.

General insights

The results presented in the previous paragraphs suggest that it is necessary, if feasible, to augment classical hedonic regression analysis with averages- and diversity measures of (structural) characteristics of neighboring properties. While the previous paragraphs suggested an array of positive and negative effects for the various measures included in the regression analyses, we now perform a simple analysis to investigate the "total effects" to get an idea about the price differentials as a result of differences in the neighborhood's housing stock composition. More specifically, we look at the distribution of the estimated average and/or diversity effects using the following simple formula:

$$N\hat{H}S_i = \hat{\gamma}\bar{X}_{ij} + \hat{\kappa}\sigma(X_{ij}) \tag{III.3}$$

where NHS_i denotes the total neighboring housing stock effect for property i in our sample, and $\hat{\gamma}\bar{X}_{ij}$ and $\hat{\kappa}\sigma(X_{ij})$ denote the average effect and diversity effect, respectively. The results are presented in table 5.[34]

Table 5: *Average, diversity and total effects*

Location controls	Effect	St. dev.	$P5$	$P95$
No dummies	Average	0.044	-0.07	0.069
	Diversity	0.029	-0.042	0.052
	Total	0.042	-0.058	0.072
Municipality	Average	0.043	-0.066	0.066
	Diversity	0.03	-0.046	0.05
	Total	0.041	-0.057	0.067

Note: all the effects were calculated using the results presented in columns (4) and (5) of table 4.

The results presented in table 5 suggest that there is a total price differential

[34] In figures B.1 and B.2 in appendix B kernel density plots of the full distribution are provided.

of approximately 13 percent due to differences in the housing stock composition when dummy variables for the municipalities are not included and 12.4 percent when dummy variables are included. The results furthermore suggest that the price differential as a result of difference in average values of neighboring properties is larger (13.9 and 13.2 percent, respectively) than the price differential as a result of differences in diversity measures (9.4 and 9.6 percent, respectively). Finally, note that both effects are negatively correlated (the correlation statistics are equal to -0.41 and -0.4 for the model with and without municipality dummy variables, respectively), since the total effects are smaller than the sum of the respective average- and diversity effects. This implies that a higher (lower) estimated average effect is typically associated with a lower (higher) estimated diversity effect and vice versa.

In a more general sense, the results presented in columns (1)-(5) are fairly stable across the different model specifications and suggest that averages and standard deviations of characteristics of neighboring properties influence the price of the property that is up for sale. This is, for example, consistent with the results documented in a strand of literature that investigates the positive externalities of (historic) landmark buildings (e.g. Ahlfeldt & Maennig, 2010; Ahlfeldt & Mastro, 2012) and finds significant effects. The current study, however, does not only consider the spillover effects of (historic) landmark buildings, but looks at the structural characteristics of neighboring properties more generally. Observe that the R^2 increases from 0.808 to 0.826 (+2.2%) in a model without municipality dummies and from 0.848 to 0.86 (+1.4%) in a model with municipality dummies. It is therefore fair to state that including information concerning neighboring properties in traditional regression models is worthwhile and can, for example, improve the performance of automatic mass appraisal systems.[35]

5.2 Robustness & extensions

Although the analyses presented in the previous section provide important insights, it remains to perform some robustness checks and extensions. In table B.3 in appendix B we present the results of a similar regression analysis, but use alternative measures of diversity. In table B.4 in appendix B the

[35] In his paper, Lindenthal (2016), also argues that including shape information that partially explains transaction prices of buildings improves the performance of automatic mass appraisal systems.

model is re-estimated using an inverse-distance spatial weights matrix W and in table B.5 in appendix B we present the estimated regression coefficients using a spatial lag model where we allow for spatial spillover effects in the dependent variable. The results presented in all these tables are all highly similar to those presented in section 5.1, which suggests that the results are not driven by the specification of the spatial weights matrix or the type of spatial spillover effects (lag versus error).

Level of aggregation

As previously mentioned, one of the major drawbacks of using the administrative data is that they do not allow us to aggregate flexibly. Nonetheless, the results presented in the previous section might be sensitive with respect to the spatial definition of neighborhood. Therefore, we perform a number of regression analyses where we use the same set of variables in the regression, but at different levels of aggregation.[36] More specifically, we have calculated the averages and standard deviations of the sizes, shapes, height and distances of buildings and plots for different levels of aggregation. The levels of aggregation considered are (1) 100 meter radius, (2) street, (3) 10 nearest neighbors, and (4) statistical sectors. The results are presented in table 6.

[36] Since the types of buildings and the year of construction is only known from the administrative data, we cannot allow these to vary flexibly and keep using the administrative data for their respective variables.

Table 6: *Robustness with respect to alternative levels of aggregation*

Category	Variable	100 meter	Street	10 NN	LLA
Avg.:	Built-on surf. (ln)	0.083***	0.0642***	0.0507***	-0.0095
		(0.0195)	(0.0183)	(0.0153)	(0.0176)
	Shape buildings	0.0204	0.0074	-0.1014*	0.0493
		(0.0805)	(0.0695)	(0.0577)	(0.1598)
	Height	0.0222***	0.0209***	0.0164***	0.0273***
		(0.0026)	(0.0023)	(0.0021)	(0.0044)
	Dist. to road	0.0068***	0.0056***	0.0048***	0.0054**
		(0.0015)	(0.0014)	(0.0013)	(0.0022)
	Plot size (ln)	-0.0212*	-0.0037	0.0164*	0.0206
		(0.0124)	(0.0115)	(0.0096)	(0.02)
	Shape plots	-0.1253***	-0.0749**	-0.0453	-0.1476**
		(0.0385)	(0.0348)	(0.0301)	(0.0701)
	Year of construction	0.0005**	0.0005*	0.0005**	0.0005
		(0.0003)	(0.0003)	(0.0003)	(0.0004)
St. dev.:	Built-on surf. (ln)	-0.0197***	-0.0104*	0.0006	-0.0107*
		(0.0061)	(0.0057)	(0.005)	(0.0059)
	Shape buildings	-0.1529**	-0.1146*	-0.0768	-0.0263
		(0.0766)	(0.0668)	(0.0597)	(0.1315)
	Height	0.0026	-0.0028	0.0022	0.0065
		(0.004)	(0.0038)	(0.0032)	(0.0062)
	Dist. to road	-0.0026***	-0.0024**	-0.0024**	-0.0014
		(0.001)	(0.001)	(0.001)	(0.001)
	Plot size (ln)	-0.0031	-0.0071	-0.0156***	0.0013
		(0.0056)	(0.0051)	(0.0047)	(0.0076)
	Shape plots	0.0376	-0.0464	0.0406*	0.1247**
		(0.0271)	(0.0387)	(0.023)	(0.0511)
D:	Buildings	-0.0106	-0.0091	-0.0197	-0.0322
		(0.0207)	(0.0205)	(0.0205)	(0.0227)
	Year of construction	-0.0237	-0.009	-0.0233	-0.0429
		(0.0327)	(0.0327)	(0.0326)	(0.0337)
Diagnostics:	ρ	0.2409***	0.2504***	0.2541***	0.2465***
		(0.0312)	(0.031)	(0.0312)	(0.0312)
	Obs.	4,276	4,270	4,222	4,281
	R-sq.	0.857	0.858	0.858	0.855
	Mun. FE	Yes	Yes	Yes	Yes

Note: Standard errors are reported in parentheses. *, ** and *** indicate $p < 0.10$, $p < 0.05$ and $p < 0.01$, respectively. Besides the reported coefficients, all the regression analyses presented furthermore include an extensive list of property- and neighborhood characteristics that is equivalent to those presented in table B.1 in appendix B.

The results presented in the different columns of table 6 are fairly similar, which suggests that the results presented earlier are robust with respect to the level of aggregation considered. Note that R^2 is lower for all specifications compared to the base model presented in column (5) of table 4. It is likely that this is due to the fact that we no longer take into account the average size of- and heterogeneity in interior space, but use the *built-on surface* which is available for all levels of aggregation in the geospatial data, as its proxy. Note furthermore that the number of statistically significant coefficients and the R^2 are higher in the first three specifications. This seems to suggest that buyers especially take into account the characteristics of neighboring properties at low levels of aggregation. This is, for example, consistent with the findings of Ahlfeldt & Mastro (2012), who show that the spillover effects of neighboring dwellings designed by the famous architect Frank Lloyd Wright (steeply) decay with distance.

Relative size effects

We also estimate a model where we allow for *relative size effects* instead of *absolute size effects*. It is, however, intuitive that both are highly correlated since an increase in the own size of the property implies a decrease in its relative size when holding the size of neighboring properties constant. In constructing the relative size measures we follow Turnbull *et al.* (2006) and allow for asymmetric size effects and potential non-linearities.[37] We then re-estimate the baseline model with dummy variables for the municipalities[38], but use relative size measures for the interior space, plot size and the year of construction. The model is again estimated using a spatial error model and the *10 nearest neighbors* spatial weights matrix. The results are presented in table 7 and figure 2.

[37]More specifically, we first construct standardized measures of "relative size" in the following way: $\tilde{k}_i^J = (k_i - \Sigma_{j \in J} k_j)/\Sigma_{j \in J} k_j$, where k, i and j denote the characteristic, property and neighborhood, respectively. We subsequently construct the variables $k: smaller$ and $k: larger$. $k: smaller$ is equal to its absolute value whenever the relative size measure k is smaller than zero. $k: larger$ is equal to its absolute value whenever the relative size measure for k is larger than zero. In a final step, we also calculate their respective squared values.

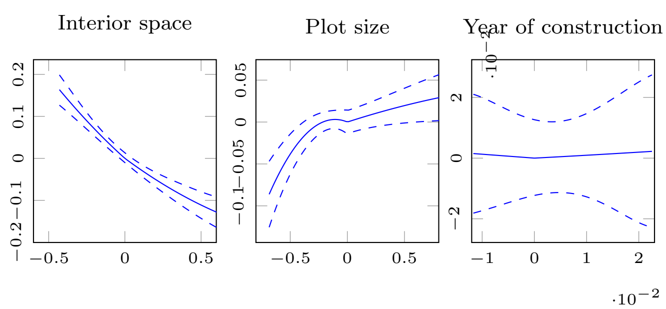
[38]As a robustness check, we have also estimated a model without dummy variables for the municipalities and the results are very similar.

Table 7: *Relative size effects*

Variable	(1)	Variable	(1)
Int.: larger	-0.268***	Plot: smaller sq.	-0.1622***
	(0.0354)		(0.0529)
Int.: larger sq.	0.0916***	YoC: larger	-1.0492
	(0.0193)		(1.0523)
Int.: smaller	0.3255***	YoC: larger sq.	58.8539*
	(0.0553)		(30.0319)
Int.: smaller sq.	0.1283	YoC: smaller	1.2622
	(0.102)		(1.5393)
Plot: larger	0.0441***	YoC: smaller sq.	-105.4014
	(0.0163)		(101.7703)
Plot: larger sq.	-0.0098***	ρ	0.214***
	(0.0036)		(0.0318)
Plot: smaller	0.0574	Obs.	4,276
	(0.0351)	R-sq.	0.861

Note: Standard errors are reported in parentheses. *, ** and *** indicate $p < 0.10$, $p < 0.05$ and $p < 0.01$, respectively. Besides the reported coefficients, all the regression analyses presented furthermore include an extensive list of property- and neighborhood characteristics that is equivalent to those presented in table B.1 in appendix B. We additionally control for municipality fixed effects.

Figure 2: *Relative size effects*



Note: the dashed lines represent the 95% confidence intervals.

Although the reported regression coefficients are now somewhat harder to interpret, the relationships presented in figure 2 show that there is a positive effect of being located in the presence of relatively larger dwellings and a negative effect associated with relatively larger plots, which is consistent with earlier findings. Note however that we no longer find a significant effect for the relative year of construction.

Heterogeneity in valuations

In all the analyses carried out so far, we did not allow for interactions between the degree of atypicality of the property and the degree of heterogeneity at the neighborhood level. It could however be the case that atypical properties are valued differently in homogeneous and heterogeneous neighborhoods. In the current section, we therefore extend the analyses presented previously by allowing for several interaction effects. More specifically, we augment the (spatial) hedonic model with the following variables:

$$\text{Abs. dev. } k_{i,j} = |\frac{k_{i,j} - \bar{k}_j}{\bar{k}_j}| \tag{III.4}$$

and:

$$\text{Int. } k_{i,j} = \text{Abs. dev. } k_{i,j} * D_{k,j} \tag{III.5}$$

where j denotes the neighborhood property i is located in, and k and D denote the characteristic and Simpson (1949)'s D, respectively. The first variable simply denotes the absolute percentage deviation of characteristic k for property i with respect to its neighborhood (denoted by j) average. When the estimated coefficient is positive (negative), this suggests that more atypical properties are more (less) expensive. The second variable captures whether properties that have a more atypical value of characteristic k are valued more (less) in neighborhoods that are characterized by a higher (lower) level of diversity in characteristic k. The results from the regression analysis are presented in table 8.

Table 8: *Heterogeneous valuations*

Variable	k	$\hat{\beta}$	$\hat{\sigma}$
Dev.	Building type	0.0098	(0.0305)
	Year of constr.	-0.01	(0.0855)
	Living area	0.0794	(0.0721)
	Plot size	-0.0513**	(0.0225)
Int.:	Building type	-0.0805	(0.0523)
	Year of constr.	-0.0125	(0.1068)
	Living area	-0.1039	(0.1277)
	Plot size	0.0796**	(0.0337)
Diagnostics:	ρ	0.2137***	(0.0321)
	Obs.	4,276	
	R-sq.	0.861	

Note: Standard errors are reported in parentheses. *, ** and *** indicate $p < 0.10$, $p < 0.05$ and $p < 0.01$, respectively. Besides the reported coefficients, all the regression analyses presented furthermore include an extensive list of property- and neighborhood characteristics that is equivalent to those presented in table B.1 in appendix B. We additionally control for municipality fixed effects. The results of the "average effects" and "diversity effects" are reported in table B.6 in appendix B. The deviation for the building type was calculated as 1 minus the share of the building type of property i.

The results presented in table 8 reveal that there is little evidence for heterogeneous valuations. We only observe that properties that have a more atypical plot size are valued negatively. Notice, however, that the magnitude of this effect decreases whenever the neighborhood is characterized by a higher degree of heterogeneity in plot sizes.

6 Concluding remarks

In this paper we started from the observation that although housing economists have frequently viewed houses as "*bundles of characteristics*" and used property attributes to explain sales prices, it has been far less common to view neighborhoods as "*bundles of properties*" and include the characteristics of neighboring properties in traditional hedonic regression analysis. Nevertheless, research has shown that beauty and aesthetic characteristics can have a significant effect on perceived community satisfaction and property values. While a large number of studies has explicitly focused on (positive) spillover effects of certain well-defined (historic) landmark buildings, we focus more generally on the composition of the housing stock at the neighborhood level in this paper. More specifically, we investigate whether the average values for- and diversity measures of certain objectively measurable characteristics, such as type, age, size and shape at the neighborhood level influence home values. Therefore, we augment a large and detailed dataset of individual transactions with administrative and geospatial data concerning the housing stock at the neighborhood level for the Flemish part of the Brussels Metropolitan Area for the period 2003-2015. The results of a (spatial) hedonic model indicate that the price of an observationally equivalent property is increasing in the average size, year of construction, roof height and distance to the road of neighboring properties, but decreases in the average size and shape of plots. Our finding that the price of a property is increasing in the average size of neighboring properties provides supporting evidence for the *tax capitalization hypothesis* of Hamilton (1976), who argues that relatively smaller houses benefit from the presence of larger houses. The results presented for the diversity measures reveal that home values are higher in neighborhoods that are characterized by low levels of diversity in (1) the types of buildings, (2) the year of construction, (3) the living area, (4) the shapes of buildings, and (5) distance to the road. Our findings indicate that differences in the housing stock composition at the neighborhood level can lead to price differentials as large as 12% across neighborhoods for an equivalent property.

While the results presented are just a single example of the use of "Big Data" in modern day econometrics, it is also intuitive that the results presented have important and potentially far-reaching consequences for policy makers, real estate professionals, and urban planners. We, for example, find that an equivalent property is worth 0.8% more in a neighborhood where the average property is 10 years younger, which suggests that the physical deterioration of neighboring properties has an impact upon the price/value of the property that is up for sale. This consequently implies that renovation subsidies for the exterior might not only have an impact upon the value and quality of renovated properties but also have spillover effects on the values of neighboring properties that should be taken into account in their cost-benefit analysis. Our results also provide backing for the often-heard realtor's advice to buy a small house in a neighborhood that is dominated by larger houses. The current paper also shows that data concerning neighboring properties that is becoming ever more available in the form of geospatial data can be used to improve mass appraisal systems.

A Appendix A: Overview of the data

A.1 Transaction data

Table A.1: *Overview of the variables in the transaction data*

Variable	Description
Transaction price	Sales price (in euro)
Date of sale	
Location	x- and y-coordinates
Type of construction	(Semi-)detached vs. terraced housing
Year of construction	
Living surface (in sq. m.)	
Plot size (in sq. m.)	
Number of bedrooms	
Number of garages	
State of property	Ready to move in, luxuriously finished
Heating type	Central heating
Heating material	Gas, wood, electricity,..
Heating elements	Radiators, underfloor heating,..
Hot water	Condensing boiler,..
Glazing	Single, double, triple,..
Basement	Storage room, wine cellar,..
Bathroom	Two/multiple bathrooms, bathtub,..
Kitchen	Well-maintained, luxuriously finished, dishwasher,..
Various	Fireplace, alarm, airconditioning,..

Table A.2: *Moran's I statistics for different variables*

Variable	I	z
Sales price (in €)	0.267808***	59.271943
Living area (in sq. m.)	0.085973***	19.051213
Plot size (in sq. m.)	0.166771***	36.998163
# Bedrooms	0.051255***	11.370069
# Garages	0.147593***	32.675321
Year of construction	0.101192***	22.415327

A.2 Administrative data housing stock composition

Table A.3: *Overview structural characteristics and their respective threshold values (January 1, 2011, source: General Administration of Patrimonial Documentation (GAPD))*

Characteristic (l)	Threshold values	# of classes
# Buildings	integer, 0-max.	infinite
# Residences	integer, 0-max.	infinite
Year of construction	1899, 1918, 1945, 1961, 1970, 1980, 1990 and 2000	9
Living area (in sq. M.)	44, 64, 104, 184, 344 and 664	7
Plot size (in sq. M.)	44, 64, 104, 184, 344, 664, 1304, 2584, 5144 and 10264	11
Built-up area (in sq. M.)	44, 64, 104 and 184	5
# Floors	1, 3 and 5	4
Cadastral revenue (in €)	499, 744, 999, 1499 and 2499	6

Note: the threshold categories indicate that the value for the respective variable has to be smaller than or equal to the threshold value. Mention that no lower and upper bounds are mentioned, although there are obviously are certainly lower bounds for almost all variables (e.g. areas, # of floors and cadastral revenue cannot be smaller than zero).

Table A.4: *Correlation statistics housing stock diversity measures*

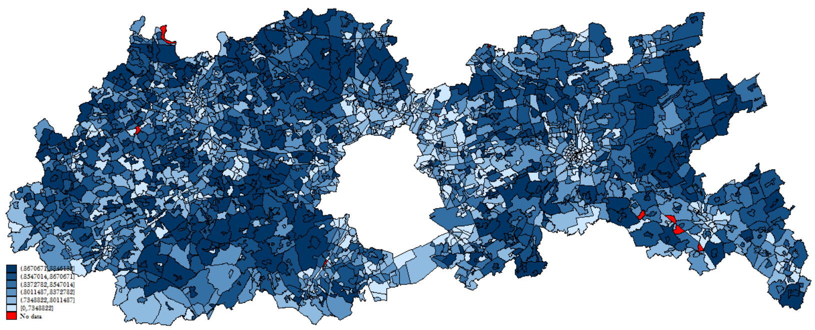
variable	Measure	*COV*	*D*	*Max*
Building types	*COV*			
	D		1	-0.958
	MAX		-0.958	1
Living area	*COV*	1	0.815	-0.656
	D	0.815	1	-0.919
	MAX	-0.656	-0.919	1
Plot size	*COV*	1	0.872	-0.867
	D	0.872	1	-0.969
	MAX	-0.867	-0.969	1
Year of constr.	*COV*	1	0.594	-0.567
	D	0.594	1	-0.95
	MAX	-0.567	-0.95	1

Table A.5: *Property characteristics own property vs. property characteristics neighboring properties*

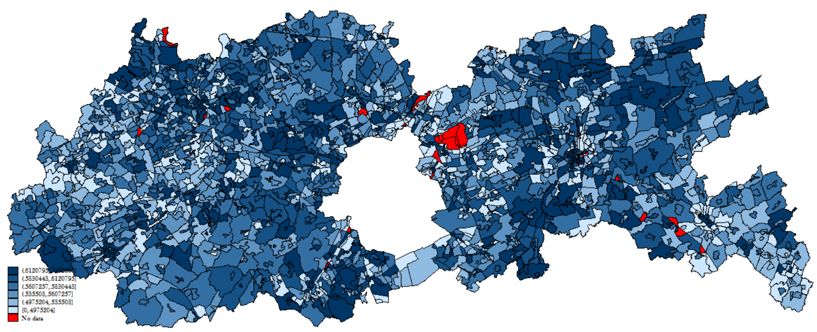
			$(k_{ij} - \bar{k}_j)$		$\|k_{ij} - \bar{k}_j\|$	
Variable	Obs.	Corr.	Avg.	St. Dev.	Avg.	St. Dev.
Living area	6,450	0.257	0.355	70.492	51.771	47.839
Plot size	6,450	0.413	221.547	990.463	437.926	915.586
Year of constr.	5,489	0.373	6.472	20.944	16.506	14.424

Figure A.1: *Simpson's diversity index D for (a) year of construction, (b) living area and (c) plot size*

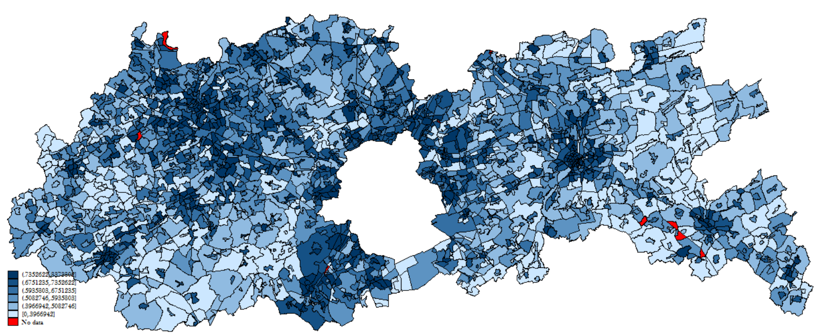
(a) Year of construction



(b) Living area



(c) Plot size



A.3 GIS-data

Description of the data

In this paper, we use information from a large bundle of shapefiles[39] that were provided by the Flemish Geographical Information Agency (FGIA)[40]. We use information from five main databases:

1. *Central Reference Address Database (CRAD):*
   The CRAD contains the street names, house numbers and information about the geographical positioning of approximately 3.4 million Flemish addresses. The data are available as a point shapefile in the coordinate reference system (CRS) *Lambert 72* (EPSG: 31370).

2. *Large-Scale Reference Database (3D-version, LRD-3D)*:
   The LRD-3D is an object oriented reference map of Flanders with precise and current information on buildings, administrative parcels, roads (and their lay-out), watercourses, railways and works of art. The database contains several polygon shapefiles in the CRS *Lambert 72* (EPSG: 31370). We focused our attention towards information on the approximately 4.9 million buildings in Flanders. For each building, we know its exact location and lay-out, surface, perimeter, and maximum roof height.

3. *Cadastral parcelling plans (CPP):*
   The CPP is an object oriented reference map of the properties as they are know by the cadastre. The data are available as a polygon shapefile in the CRS *Lambert 72* (EPSG: 31370) and contain information on the location, lay-out, surface and perimeter of the approximately 4.7 million Flemish administrative parcels.

4. *Mid-scale Reference Database Roads (MRD-Roads):*
   The MRD-Roads is a mid-scale reference database of the roads in Flanders. It contains the location and lay-out of all Flemish roads together with their matching attribute data. The data are available as a polygon shapefile in the CRS *Lambert 72* (EPSG: 31370).

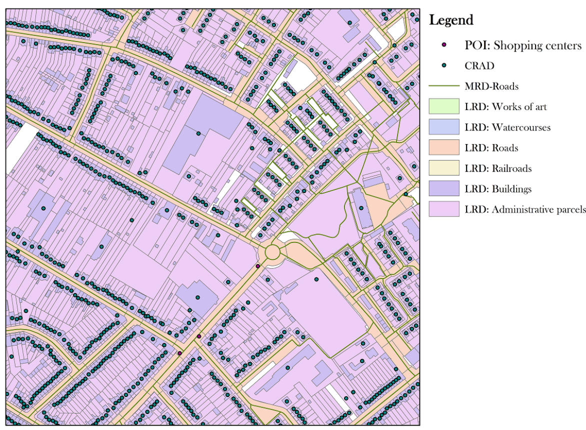
[39] The shapefile is a popular geospatial vector data format for geographic information system (GIS) software and was/is developed by Esri as a (mostly) open specification for data interoperability among Esri and other GIS software product.

[40] The Flemish Geographical Information Agency was founded in 2006 and is responsible for provision of (Flemish) geographical data to government agencies, businesses and citizens.

5. *Points of interest (POIs):*
   the POI database contains the exact location of various POIs, such as bus stops, train station, highway entries/exits, grocery stores, shopping centers, schools, town halls, and so on. The data are available as a point shapefile in the CRS *Lambert 72* (EPSG: 31370).

A geographical overview of the different databases is presented in figure A.2.

Figure A.2: *Visual representation databases Flemish Geographical Information Agency (FGIA)*



Source maps: FGIA

Managing the data

In a first step, we merge the transaction data ($\approx$ 6.450 transactions) with the data from the Central Reference Address Database. For 6.405 transactions we were able to merge the address contained in the transaction database to the address information in the CRAB.

In a second step, the different spatial databases provided by the FGIA are merged using the appropriate spatial join procedures in ArcGIS. Firstly, the CRAD is merged with the administrative parcels. Secondly, the centroids of the buildings in the LRD-3D are spatially joined to the administrative parcels. By calculating the distance between the facade points of the different buildings to the different road segments in the MRD-Roads database, we can also determine the minimum distance to the road for every building in Flanders. Since the attribute data in the MRD-Roads also contains information on the type of road, we also know the type of road for every building in Flanders. Using the location information on various amenities (bus stops, train stations, highway entries/exits, grocery stores, shopping centers, and so on) we can furthermore calculate the (crow-fly) distance to these amenities and the number within a certain radius. To summarize, for every address in Flanders, we know (1) the surface (in m^2) and perimeter (in m) of the administrative parcel on which it is located, (2) the type (*main building* versus *outhouse*), surface (in m^2), perimeter (in m), maximum roof height (in m) and minimum distance to the road (in m) for every that is located on that administrative parcel and (3) the type of road (*highway*, *primary*, *secondary*, *tertiary*, *residential*) along which the address is located.

In a third step, we calculate some additional variables that provide information concerning the shape of the building/parcel using its respective surface and perimeter. The following shape measure is employed:

$$\text{Shape} = \frac{\text{Perimeter (in } m\text{)/Surface (in } m^2\text{)}}{4/\sqrt{\text{Surface (in } m^2\text{)}}} \tag{III.6}$$

where the numerator represents a normalization, since the denominator for an equally shaped parcel/building is increasing in its surface. For a perfectly square plot/building, this ratio reduces to $(4s/s^2)/(4/s) = 1$, where s is the side of the object. The shape variable increases when the plot/building is more elongated and/or irregularly shaped.

Descriptive statistics

Table A.6: *Descriptive statistics properties using GIS-data*

Variable	Obs.	Avg.	St. Dev.	P5	P95
Plot size (in sq. m.)	5,668	653.7	794.5	96.57	1,876
Perimeter plot (in m.)	5,668	116.8	80.93	46.6	231.4
Shape plot	5,668	1.27	0.26	1.008	1.757
# Buildings	5,668	1.582	0.864	1	3
Total built up area (in sq. m.)	5,668	146.8	99.3	63.1	284
Avg. shape main building	5,668	1.155	0.143	1.008	1.445
Max. height main building	5,668	8.872	2.058	5.48	11.89
Min. dist. to road	5,668	4.68	7.11	0	12.91
Perc. built upon	5,668	0.355	0.298	0.085	0.974

Table A.7: *Descriptive statistics neighboring properties constructed from geospatial data (street)*

Type	Char.	Statistic	Obs.	Avg.	St. Dev.	P5	P95
Building	Dist. to road	Avg.	6,041	5.149	5.467	0	12.01
		St. dev.	6,006	3.742	4.126	0	11.11
	Height	Avg.	6,041	8.846	1.535	6.607	11.36
		St. dev.	6,006	1.494	0.668	0.531	2.575
	Number	Obs.	6,041	20.44	12.5	5	45
	Perimeter	Avg.	6,041	52.65	10.97	38.25	70.14
		St. dev.	6,006	15.47	13.61	4.467	38.36
	Shape	Avg.	6,041	1.153	0.073	1.052	1.278
		St. dev.	6,006	0.119	0.06	0.035	0.22
	Surface	Avg.	6,041	139.7	67.28	78.5	220.7
		St. dev.	6,006	77.7	130.9	15.97	227.7
Plot	Number	Obs.	6,041	19.91	11.98	5	43
	Perimeter	Avg.	6,041	119.7	41.77	69.07	187.2
		St. dev.	6,011	43.75	34.47	8.671	93.55
	Shape	Avg.	6,041	1.257	0.142	1.058	1.505
		St. dev.	6,011	0.177	0.099	0.039	0.352
	Surface	Avg.	6,041	718.6	684.4	191.5	1,593
		St. dev.	6,011	501.3	596.3	69.4	1,402

Table A.8: *Descriptive statistics neighboring properties constructed from geospatial data (10 nearest neighbors)*

Type	Char.	Statistic	Obs.	Avg.	St. Dev.	P5	P95
Building	Dist. to road	Avg.	6,107	4.743	4.478	0	11.47
		St. dev.	6,100	3.1	3.818	0	9.914
	Height	Avg.	6,107	8.839	1.587	6.571	11.47
		St. dev.	6,100	1.422	0.756	0.312	2.741
	Number	Obs.	6,107	9.197	1.947	6	12
	Perimeter	Avg.	6,107	50.96	10.3	36.61	67.97
		St. dev.	6,100	12.98	10.4	3.502	30.89
	Shape	Avg.	6,107	1.15	0.079	1.046	1.297
		St. dev.	6,100	0.111	0.062	0.029	0.223
	Surface	Avg.	6,107	129.4	50.96	71.84	207.4
		St. dev.	6,100	56.06	75.01	11.42	150.5
Plot	Number	Obs.	6,109	11	0	11	11
	Perimeter	Avg.	6,109	111.7	38.65	59.05	184.1
		St. dev.	6,109	43.27	38.88	7.384	99.13
	Shape	Avg.	6,109	1.27	0.149	1.076	1.549
		St. dev.	6,109	0.205	0.143	0.05	0.445
	Surface	Avg.	6,109	622.9	467.7	145.5	1,558
		St. dev.	6,109	450.5	565.1	48.57	1,473

Table A.9: *Descriptive statistics neighboring properties constructed from geospatial data (statistical sectors)*

Type	Char.	Statistic	Obs.	Avg.	St. Dev.	P5	P95
Building	Dist. to road	Avg.	6,273	5.909	3.625	1.04	12.81
		St. dev.	6,273	7.982	5.845	2.893	18.99
	Height	Avg.	6,273	8.742	1.188	7.227	10.84
		St. dev.	6,273	2.055	0.484	1.489	2.872
	Number	Obs.	6,273	427.8	256.9	100	873
	Perimeter	Avg.	6,273	54.9	9.903	42.89	67.55
		St. dev.	6,273	27.98	20.54	11.6	55.46
	Shape	Avg.	6,273	1.153	0.047	1.078	1.229
		St. dev.	6,273	0.141	0.038	0.081	0.201
	Surface	Avg.	6,273	165.3	126.1	98.52	240
		St. dev.	6,273	252.4	457.1	50.99	645.2
Plot	Number	Obs.	6,273	661.6	473.6	183	1,560
	Perimeter	Avg.	6,273	145.5	57.51	81.75	269.8
		St. dev.	6,273	100.5	51.16	47.36	201.3
	Shape	Avg.	6,273	1.279	0.07	1.175	1.405
		St. dev.	6,273	0.298	0.073	0.203	0.426
	Surface	Avg.	6,273	1,326	1,273	299.7	4,274
		St. dev.	6,273	2,428	2,962	502.9	7,500

A.4 Other data

Table A.10: *Descriptive statistics data Census 2011 & taxable income statistics (ADS)*

Variable	Obs.	Avg.	St. Dev.	P5	P95.
Area (in sq. km.)	2,240	1.081	1.536	0.12	3.964
# Inhabitants	2,213	598.5	557.8	33	1,719
Inh./sq. km.	2,213	1,434	1,779	32.46	5,161
% Belgian	2,213	94.21	8.337	77.52	100
% Women	2,213	50.69	5.023	45.87	55.42
% Married	2,213	43.75	8.67	31.05	54.55
% Divorced	2,213	6.785	4.314	2.83	11.23
Avg.: age	2,213	41.97	4.183	36.49	48.2
Max: age	2,213	0.302	0.07	0.245	0.387
D: age	2,213	0.779	0.066	0.738	0.806
St. dev.: age	2,213	21.64	2.252	19.31	23.76
COV: age	2,202	1.938	0.314	1.735	2.202
Avg.: education	2,211	4.204	0.43	3.626	4.946
Max: education	2,211	0.373	0.098	0.28	0.554
D: education	2,211	0.727	0.081	0.617	0.78
% Student	2,213	7.891	3.212	3.704	12.55
% Retired	2,213	1.682	2.426	0	3.627
% Inactive labor m.	2,213	53.56	7.365	45.45	64.75
% Unemployed	2,213	1.682	2.426	0	3.627
# Tax returns	2,212	337.6	323.2	18	991
Avg. tax. inc.	1,975	34,282	6,987	25,193	46,519
Median tax. inc.	2,091	25,573	3,875	19,713	31,840
Interquartile coefficient (IC)	2,091	111.1	24.7	80	152
# Residences	2,212	244.8	238.8	12	730
% Owner-occupied	2,212	78.01	16.71	45.72	95
% Rental	2,212	21.1	16.39	4.348	52.73
Avg.: # rooms	2,211	5.807	0.756	4.621	6.846
Max: # rooms	2,211	0.317	0.104	0.214	0.488
D: # rooms	2,211	0.783	0.095	0.667	0.856
St. dev.: # rooms	2,211	1.44	0.282	1.085	1.85
COV: # rooms	2,192	4.143	1.092	2.674	5.455

B Appendix B: Regression results

Table B.1: *Results baseline model without diversity measures*

Category	Variable	(1)	(2)	Category	Variable	(1)	(2)
Type of construction:	Semi-detached	0.0188***	0.0187***		Electric stove	-0.0054	-0.0016
	Detached	0.0774***	0.0776***		Fully fitted	0.0011	0
Surface:	Ln(living area)	0.2138***	0.2277***		Gas stove	-0.0125*	-0.0103
	Ln(Plot size)	0.1297***	0.1248***		Hood	0.0091	0.0177***
# Bedrooms:	1	-0.071***	-0.075***		Oven	0.0083	0.0071
	3	0.0558***	0.0526***		Refrigerator	0.0103**	0.0123**
	4	0.0856***	0.083***		Well-maintained	0.0022	0.0016
	5	0.134***	0.1312***	Bathroom:	Double sink	0.0153**	0.0147**
	6	0.1495***	0.1378***		Ind. per bedroom	0.0862**	0.0712*
	7	0.3766***	0.3836***		Multiple	0.0681***	0.0698***
# Garages:	1	0.0257***	0.028***		Separate toilet	-0.0008	0.0004
	2	0.0436***	0.0419***		Two	0.0395***	0.0405***
	3	0.0446**	0.0409**	Basement:	Total basement	0.0279***	0.0292***
	4	0.0849**	0.0949**		Wine cellar	0.0737***	0.0658***
Age:	Linear	-0.005***	-0.0045***	Environment:	Forest/parcs	0.0467***	0.0291**
	Squared	0***	0***		Free sight	0.0045	-0.0022
State:	Luxuriously finished	0.087***	0.0864***	Various:	Alarm	0.0356***	0.0355***
	Major refurbishment nec.	-0.0355***	-0.037***		Automatic garage door	0.002	0.0015
	Minor refurbishment nec.	-0.0052	-0.0067		Fireplace	0.0239***	0.0195***
	Ready to move in	0.0773***	0.0791***	Census:	Ln(Med. tax. inc.)	0.1034***	0.0748***
	Total reconstruction nec.	-0.1934***	-0.1872***		Ln(pop. density)	0.0082***	0.0077**
Heating type:	Central heating	0.0599***	0.0555***	Location:	Dist. to city hall	-0.0015	-0.0001
Heating material:	Coal	0.0054	-0.0067		Dist. to highway	0.0058**	0.0011
	Electricity	0.0407***	0.044***		Leuven region	0.2069***	0
	Gas	0.0582***	0.0597***		Travel time Brussels	-0.0153***	-0.0066***
	Oil fuel	0.054***	0.0558***		Travel time Leuven	-0.0116***	-0.0042
Heating elements:	Accumulation	0.0364**	0.03*		# Bars/pubs 1km	0.0006	0.0012**
	Convectors	0.0215**	0.0211**		# Grocery st. 1km	0.0079**	0.0065**
	Radiators	0.0237***	0.0243***		# Schools 1km	-0.0014	0.0011
	Underfloor heating	0.0679***	0.0697***	Year of sale:	2006	0.0763***	0.078***
Hot water:	Condensing boiler	0.0515***	0.043***		2007	0.1343***	0.1352***
	Electric boiler	-0.0058	-0.0043		2008	0.1325***	0.1319***
	Flowthrough system	0.0072	0.007		2009	0.1386***	0.1393***
	Gas boiler	0.0016	0.0011		2010	0.1708***	0.1713***
	Gasgeyser	-0.0135*	-0.0109		2011	0.2107***	0.2147***
Glazing:	Double	0.0011	0		2012	0.2141***	0.2186***
	Single	-0.0392***	-0.039***		2013	0.1998***	0.2061***
	Triple	0.0346	0.0454		2014	0.2136***	0.2164***
Kitchen:	All comfort	0.0443***	0.0438***		2015	0.222***	0.2238***
	Breakfast area	0.0126**	0.0054	Diagnostics:	ρ	0.5626***	0.2796***
	Ceramic stove	0.0093	0.0117*		Obs.	4,276	4,276
	Dishwasher	0.0288***	0.0294***		R-sq.	0.808	0.848
	Double sink	0.0064	0.0077		Location controls:	None	Municipality

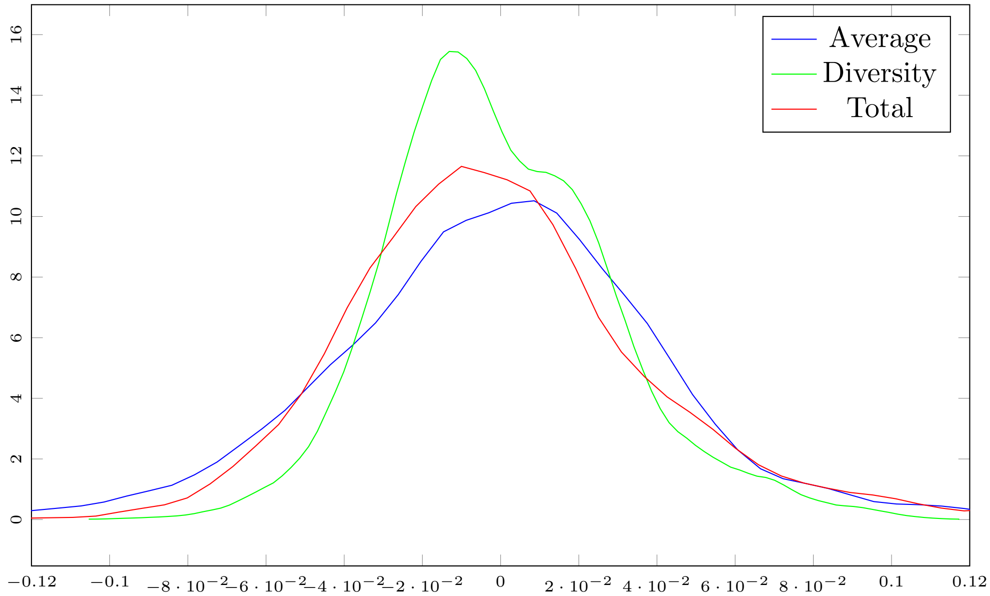
Note: Standard errors are reported in parentheses. *, ** and *** indicate $p < 0.10$, $p < 0.05$ and $p < 0.01$, respectively.

Table B.2: *Baseline results: additional features property retrieved from geospatial data*

Category	Variable	(1)	(2)	(3)	(4)	(5)
Road:	Dist. to road	0.0023***	0.0023***	0.0022***	0.0012**	0.0013**
		(0.0005)	(0.0005)	(0.0005)	(0.0006)	(0.0006)
	Primary	-0.0658***	-0.0616***	-0.0639***	-0.0673***	-0.0544***
		(0.0123)	(0.0124)	(0.0122)	(0.0123)	(0.0119)
	Secondary	-0.0401***	-0.0357***	-0.0381***	-0.0367***	-0.0384***
		(0.0109)	(0.011)	(0.0108)	(0.0108)	(0.0105)
	Tertiary	-0.0184***	-0.0135**	-0.017***	-0.0134**	-0.0105*
		(0.006)	(0.0061)	(0.006)	(0.0061)	(0.0059)
Shape:	Building	-0.0356**	-0.0316*	-0.036**	-0.0209	-0.0206
		(0.0176)	(0.0177)	(0.0175)	(0.0183)	(0.018)
	Plot	-0.0758***	-0.075***	-0.0738***	-0.0621***	-0.0625***
		(0.0101)	(0.0102)	(0.0101)	(0.0106)	(0.0105)

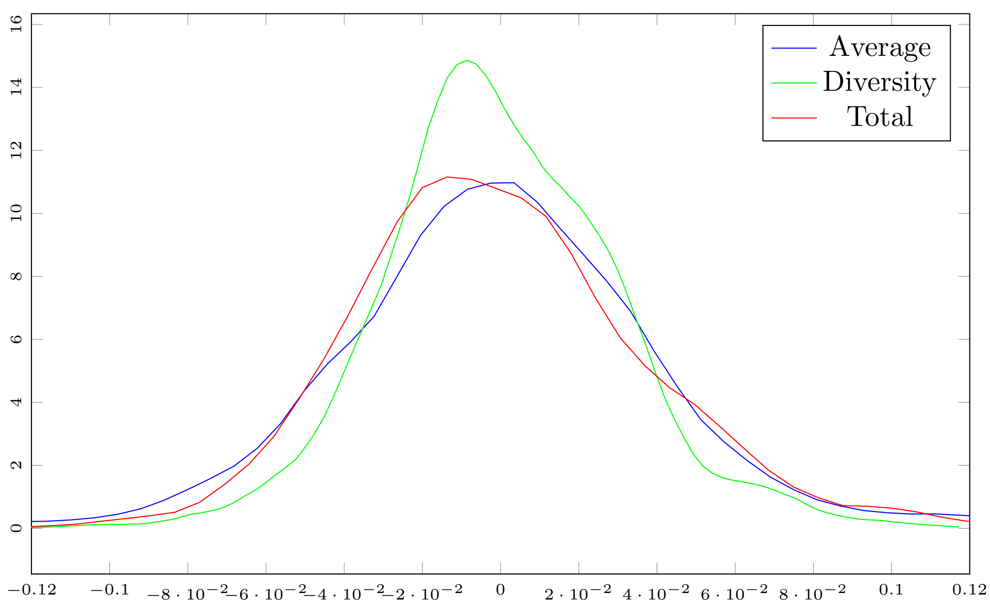
Note: Standard errors are reported in parentheses. *, ** and *** indicate $p < 0.10$, $p < 0.05$ and $p < 0.01$, respectively.

Figure B.1: *Kernel density plots average-, diversity- and total housing stock composition effects*



Note: all the results were obtained using the *kdensity*-procedure in Stata 14.

Figure B.2: *Kernel density plots average-, diversity- and total housing stock composition effects (2)*



Note: all the results were obtained using the *kdensity*-procedure in Stata 14.

Table B.3: *Robustness with respect to alternative measures of diversity*

Category	Variable	(1)	(2)	(3)	(4)
Ln(avg.)	Living area	0.317***	0.3506***	0.2781***	0.2734***
		(0.0446)	(0.0417)	(0.0332)	(0.0305)
	Plot size	-0.0567***	-0.0765***	-0.0209	-0.0328*
		(0.0217)	(0.0206)	(0.0211)	(0.0196)
Avg.:	Year of construction	0.0012***	0.0008**	0.001***	0.0006**
		(0.0004)	(0.0004)	(0.0003)	(0.0003)
	Shape buildings	0.0305	0.0883	0.0306	0.0883
		(0.078)	(0.0758)	(0.0776)	(0.0755)
	Shape plots	-0.1538***	-0.1331***	-0.1521***	-0.1336***
		(0.0382)	(0.0363)	(0.0383)	(0.0364)
	Height	0.0145***	0.0165***	0.0146***	0.0166***
		(0.0028)	(0.0027)	(0.0028)	(0.0027)
	Dist. to road	0.0057***	0.0059***	0.0058***	0.006***
		(0.0015)	(0.0014)	(0.0015)	(0.0014)
D:	Buildings	-0.0693***	-0.0236		
		(0.0259)	(0.0234)		
Ln(st. dev.)	Living area	-0.0461*	-0.0796***		
		(0.0236)	(0.0221)		
	Plot size	0.0354*	0.0381**		
		(0.02)	(0.018)		
St. dev.:	Year of construction	-0.001*	-0.0008		
		(0.0006)	(0.0006)		
	Shape buildings	-0.1821**	-0.2009***	-0.1838**	-0.201***
		(0.0761)	(0.0734)	(0.0758)	(0.0732)
	Shape plots	0.0556**	0.0388	0.0569**	0.0395
		(0.0272)	(0.0261)	(0.0272)	(0.026)
	Height	-0.0029	-0.0005	-0.0025	-0.0001
		(0.004)	(0.0038)	(0.004)	(0.0038)
	Dist. to road	-0.0019*	-0.0027***	-0.0019*	-0.0027***
		(0.001)	(0.001)	(0.001)	(0.001)
Max:	Buildings			0.0613***	0.0314
				(0.0236)	(0.0212)
	Year of construction			0.063**	0.0425*
				(0.0282)	(0.0257)
	Living area			0.101**	0.1219***
				(0.0397)	(0.0363)
	Plot size			-0.0404	-0.069**
				(0.0312)	(0.028)
Diagnostics:	ρ	0.5284***	0.2119***	0.5309***	0.2187***
		(0.0194)	(0.0323)	(0.0192)	(0.0317)
	Obs.	4,276	4,276	4,276	4,276
	R-sq.	0.826	0.86	0.826	0.86
	Mun. FE	No	Yes	No	Yes

Note: Standard errors are reported in parentheses. *, ** and *** indicate $p < 0.10$, $p < 0.05$ and $p < 0.01$, respectively.

Table B.4: *Robustness with respect to the spatial weights matrix*

Category	Variable	(1)	(2)	Category	Variable	(1)	(2)
Road:	Dist. to road	0.0012**	0.0012**			(0.0029)	(0.0027)
		(0.0006)	(0.0006)		Dist. to road	0.0058***	0.006***
	Primary	-0.0631***	-0.0518***			(0.0015)	(0.0014)
		(0.0127)	(0.0121)	D:	Buildings	-0.0572**	-0.0318
	Secondary	-0.0306***	-0.0353***			(0.0251)	(0.0231)
		(0.0113)	(0.0107)		Year of constr.	-0.0893**	-0.0562
	Tertiary	-0.0145**	-0.0109*			(0.0392)	(0.0364)
		(0.0062)	(0.006)		Living area	-0.0902*	-0.173***
Shape:	Building	-0.0164	-0.0204			(0.0503)	(0.047)
		(0.0184)	(0.018)		Plot size	0.0564	0.0856**
	Plot	-0.0639***	-0.0634***			(0.037)	(0.0336)
		(0.0108)	(0.0105)	St. dev.:	Shape buildings	-0.2002***	-0.213***
Ln(avg.)	Living area	0.2677***	0.274***			(0.0773)	(0.0738)
		(0.0333)	(0.0308)		Shape plots	0.047*	0.0374
	Plot size	-0.0109	-0.0262			(0.0275)	(0.0262)
		(0.0206)	(0.0196)		Height	-0.0008	0.0007
Avg.:	Year of constr.	0.0009**	0.0008**			(0.0041)	(0.0039)
		(0.0003)	(0.0003)		Dist. to road	-0.0015	-0.0026***
	Shape buildings	0.009	0.0912			(0.001)	(0.001)
		(0.079)	(0.0762)	Diagn.:	ρ	0.522***	0.2461***
	Shape plots	-0.1433***	-0.132***			(0.0178)	(0.0292)
		(0.0386)	(0.0365)		Obs.	4,276	4,276
	Height	0.016***	0.0176***		R-sq.	0.827	0.86
					Mun. FE	No	Yes

Note: Standard errors are reported in parentheses. *, ** and *** indicate $p < 0.10$, $p < 0.05$ and $p < 0.01$, respectively.

Table B.5: *SAR model*

Category	Variable	(1)	(2)	Category	Variable	(1)	(2)
Road:	Dist. to road	0.0012*	0.0013**			(0.0027)	(0.0026)
		(0.0006)	(0.0006)		Dist. to road	0.0053***	0.0057***
	Primary	-0.0647***	-0.0513***			(0.0015)	(0.0014)
		(0.0121)	(0.0117)	D:	Buildings	-0.0585***	-0.0312
	Secondary	-0.03***	-0.0366***			(0.021)	(0.0214)
		(0.0106)	(0.0103)		Year of constr.	-0.0532	-0.0303
	Tertiary	-0.012**	-0.0093			(0.0356)	(0.0347)
		(0.006)	(0.0058)		Living area	-0.1054**	-0.1778***
Shape:	Building	-0.0179	-0.02			(0.0434)	(0.044)
		(0.019)	(0.0181)		Plot size	0.0917***	0.089***
	Plot	-0.0556***	-0.0611***			(0.0305)	(0.0309)
		(0.0111)	(0.0105)	St.dev.:	Shape buildings	-0.2098***	-0.2193***
Ln(avg.)	Living area	0.2303***	0.2494***			(0.0756)	(0.0724)
		(0.0289)	(0.0294)		Shape plots	0.0459*	0.0376
	Plot size	-0.0649***	-0.0479***			(0.0266)	(0.0256)
		(0.0176)	(0.0186)		Height	-0.0037	-0.0003
Avg.:	Year of constr.	0.0013***	0.0009***			(0.004)	(0.0038)
		(0.0003)	(0.0003)		Dist. to road	-0.0007	-0.0024**
	Shape buildings	0.0106	0.0954			(0.001)	(0.001)
		(0.0762)	(0.0748)	Diagn.:	ρ	0.002	0.003
	Shape plots	-0.1104***	-0.1228***			(0.007)	(0.0066)
		(0.0368)	(0.0355)		Obs.	4,276	4,276
	Height	0.015***	0.0171***		Mun. FE	No	Yes

Note: Standard errors are reported in parentheses. *, ** and *** indicate $p < 0.10$, $p < 0.05$ and $p < 0.01$, respectively.

Table B.6: *Average effects and diversity effects under heterogeneous valuations*

Variable	k	$\hat{\beta}$	$\hat{\sigma}$
Ln(avg.):	Living area)	0.2961***	(0.031)
	Plot size)	-0.0328*	(0.0199)
Avg.:	Year of construction	0.0007**	(0.0003)
	Max. roof height	0.0172***	(0.0027)
	Dist. to road	0.0062***	(0.0014)
	Shape building	0.0932	(0.0757)
	Shape plot	-0.1288***	(0.0364)
D:	Buildings	0.0182	(0.0324)
	Year of constr.	-0.055	(0.0423)
	Living area	-0.1536***	(0.0552)
	Plot size	0.069*	(0.037)
St. dev.:	Max. roof height	0.0007	(0.0039)
	Shape building	-0.2059***	(0.0733)
	Dist. to road	-0.0026***	(0.001)
	Shape plot	0.0422	(0.026)

Note: Standard errors are reported in parentheses. *, ** and *** indicate $p < 0.10$, $p < 0.05$ and $p < 0.01$, respectively. Besides the reported coefficients, all the regression analyses presented furthermore include an extensive list of property- and neighborhood characteristics that is equivalent to those presented in table B.1 in appendix B. We additionally control for municipality fixed effects.

Chapter IV

Intermediation in Markets with Buyer and Seller Selection: Theory and an Application to Real Estate Brokerage

1 Introduction

One-to-one matching markets are often characterized by a dual structure.[1] On the one hand, there is an industry of intermediaries who guide the matching process between market participants and they usually charge a significant fee in return for this service. On the other hand, there is an outside market in which market participants independently search for a trading partner and incur their own search costs. For example, in real estate markets there is typically a real estate brokerage industry and a "for-sale-by-owner" market. In labor markets, firms can utilize the services of a recruitment agency or can internally organize the process of hiring a new employee. Importantly, different types of participants in these markets self-select across the different matching channels. They do so, not only based on their own preferences, but also based on their expectations about which type of trading partner they will eventually meet. So, the participation decision of each market participant entails an externality for the participants on the other side of

[1]This chapter is based upon joint work with Bert Willekens, Maarten Goos and Erik Buyst. We would like to thank Liran Einav, Jon Levin, Andras Niedermayer, Glen Weyl, Jan Rouwendal, Jan Mutl, Frank Verboven, Jo Van Biesebroeck, Patrick Van Cayseele and seminar and conference participants at the University of Leuven, Stanford University, Bocconi University (EARIE 2014), University of Alicante (ENHR 2014) and University of Reading (AREUEA 2014) for useful comments and suggestions.

the market and it is important to recognize that these externalities might influence the optimal pricing behavior of the intermediaries. For example, real estate brokers might have the incentive to charge a relatively high service fee to attract only the sellers most eager to sell their property, as these sellers are the ones buyers prefer to trade with. Or, a recruiter might charge a high placement fee in order to attract only high-level job vacancies and candidates.

This paper investigates what the socially optimal size and market structure is of an industry of private intermediaries, taking into account that the service fee charged by the intermediaries influences which types of participants are selected into the intermediary market. It is important to address this question, given that the incentives of private intermediaries that maximize profits are not necessarily aligned with those of a social planner. Not only because the intermediaries might possess market power, which can certainly be the case due to the informational advantages they inherently have over the market participants, but also because they might not properly internalize the externalities present in the market when pricing their services. To address the issue, a general model of imperfect competition among intermediaries is presented and the private market outcome is compared to the socially optimal outcome. Subsequently, the model is empirically applied to the Belgian real estate brokerage industry which provides an adequate setting to test our model and for which we have unique data. Real estate is a particularly interesting application because different types of buyers and sellers typically decide to hire the services of a real estate broker compared to those who trade in the for-sale-by-owner market.[2]

In the model, interactions among market participants - referred to as buyers and sellers - and the intermediaries that operate in the market - referred to as brokers - occur in four stages.[3] In the first stage, brokers can freely enter the market as long as they expect it is profitable to do so. In the second stage, the brokers that entered the market imperfectly compete to attract buyers and sellers by announcing their service fee charged to either buyers or

[2]Hendel, Nevo and Ortalo-Magné (2009), for example, provide evidence that less patient sellers and more patient buyers (to avoid patient sellers) tend to trade in the brokerage market compared to those who trade in the for-sale-by-owner market.

[3]Equivalently, for a labor market one can think of the sellers as workers, of the buyers as firms and of the brokers as recruiters. The price of the traded "good" is then the wage of the worker.

sellers. The service fee possibly consists of a flat fee and a fee proportional to the price of the traded good and is only paid conditional on a successful transaction. In the third stage, buyers and sellers enter the intermediary market when their expected utility of participating is greater than when participating through the outside market. In the final stage, buyers and sellers that participate in the intermediary market are randomly assigned to one another and the sales price of the traded good is determined by a Nash bargain between the buyer and the seller. The assumptions in the final stage are imposed to capture the intuition that market participants care about the characteristics of their trading partner - the price a seller receives and a buyer pays depends on the reservation value of their trading partner - and that due to information imperfections there is uncertainty about which trading partner they will meet when deciding on market participation.

A first important result derived from this setting is that a social planner always charges an intermediary service fee above the per-match cost of serving buyers and sellers. The planner internalizes the externality that buyers dislike high reservation price sellers and sellers dislike low valuation buyers, which are excluded from the intermediary market by charging a relatively high service fee. It follows that some market power attributed to intermediaries is justified when they compete in a private market. The monopoly (or collusive) service fee, however, always exceeds the socially optimal fee. So, there exists an inverse u-shaped relationship between private broker market power and social value created by the intermediary industry. A second finding is that the private market outcome is generally characterized by an excessive number of intermediaries that operate in the market compared to what is socially optimal. That is, when a novel entrant steals business from incumbent brokers, valuable resources are wasted by brokers inefficiently competing to realize the same number of transactions that could also be established with fewer brokers operating in the market. Furthermore, this entry distortion is more severe when brokers possess more market power in pricing their services. So, combined with the result that some broker market power is justified to properly internalize the participation externalities of buyers and sellers, the model has nonstandard policy implications.[4]

[4]Note that it is important to also account for the entry distortion on top of the distortion in the service fee, as there are typically little barriers for novel intermediaries to enter and operate in matching markets. Hsieh and Moretti (2003) and Barwick and Pathak (2015), for example, provide evidence that the entry distortion is quantitatively important in the US real estate brokerage industry.

In comparison to the private free entry equilibrium, the welfare effects are derived when a social planner optimally regulates broker service fees, broker entry or both. When regulating both, all market distortions can be eliminated by setting the service fee such that the participation externalities of buyers and sellers are properly internalized and by minimizing the number of brokers that operate in the market. However, the welfare gains and redistributive effects of only regulating either broker service fees or broker entry are ambiguous. They depend on the underlying parameters of seller supply and buyer demand and the structure of broker costs. The model outcomes are therefore further illustrated for realistic calibrated parameter values using data from the Belgian real estate brokerage industry.[5]

The empirical results suggest that the observed average commission rate of 4.3% charged by brokers is below the socially optimal commission rate, which ranges from 5.1% to 24% for the estimated range of feasible values for the parameters of seller supply and buyer demand. This implies that the externalities present in the market are insufficiently internalized and it would be welfare improving to exclude more buyers and sellers. For the most inelastic bound on estimated supply and demand elasticities, the welfare counterfactuals suggest that a welfare gain of 19% could be established when regulating both service fees and market entry of brokers. The outcome of a social planner that regulates broker entry and allows brokers to privately compete in pricing their services is calculated to generate a welfare gain of 18%, while regulating service fees and allowing for free broker entry only implies a welfare gain of 5%. For the most elastic bound on the supply and demand elasticities, however, regulating both service fees and entry implies a welfare gain of 71%, only regulating entry results in a gain of 40% and only regulating service fees in a gain of 52%. These results suggest that it is more effective to regulate broker entry for relatively inelastic and to regulate service fees for relatively elastic supply and demand, respectively. It should also be noted, however, that all the welfare gains are gains in con-

[5] The institutional setting of the Belgian real estate market is particularly interesting to apply the model because there are no significant barriers for new brokers to enter the market and there are no institutional restrictions for brokers to compete in pricing their services. It is therefore sensible to construct welfare counterfactuals in which both the pricing and entry behavior of brokers is affected by policy interventions. This contrasts with the US, for example, where commission rates charged by real estate brokers typically show little variation, which suggests a lack of competition among brokers in pricing their services (e.g. Hsieh and Moretti 2003).

sumer surplus attributed to buyers and sellers when regulating service fees, while regulating broker entry implies a loss in consumer surplus, which is compensated by a gain in broker profits.

In the literature, there is an extensive strand of research that investigates the role of intermediaries to facilitate market transactions - see, for example, Rubinstein and Wolinsky (1987), Biglaiser (1993) and Yavas (1994) for seminal contributions. Spulber (1999) provides a unified perspective on the different views on firm intermediation in the early literature, in which intermediaries usually play the role of market clearing entities and the externalities induced by the participation decision of different buyer and seller types emphasized here play no role. More recently, Niedermayer and Shneyerov (2014) and Loertscher and Niedermayer (2015) explore how an optimal market clearing mechanism can be implemented by intermediaries that charge a service fee instead of directly setting bid-ask spreads.

Evaluating the optimal pricing behavior of platform businesses in the presence externalities across different groups of market participants has been the topic of interest in the so-called two-sided markets literature - e.g. Rochet and Tirole (2003), Armstrong (2006) and Weyl (2010). In this literature it is typically assumed, however, that only the size and not the composition of one group of market participants affects utility of another group. In our setting, it is precisely the changed composition of user types when more or less buyers and sellers participate in the intermediary market that drives the results. Damiano and Li (2007, 2008) analyze a similar composition effect. Although their setup is quite different from ours - e.g. they allow for complementarities in the match value function and analyze duopoly competition among endogenously differentiated platforms - Damiano and Li (2008) establish a similar result that the market outcome under duopoly can be less efficient than the monopoly outcome. The basic intuition for this result is the same as in our setting. Since market participants care about the quality of their trading partner, it can be socially efficient to exclude some participants from the market.[6]

Our work also relates to recent research on competition among service providers in markets where consumer selection plays an important role, like insurance

[6]Gomes and Pavan (2015) build on Damiano and Li (2007) to investigate optimal matching mechanisms in many-to-many matching settings.

and credit markets - e.g. Einav and Finkelstein (2011), Veiga and Weyl (2015) and Mahoney and Weyl (2015). In particular, Mahoney and Weyl (2015) demonstrate that an inverse u-shaped relationship exists between competition and welfare in markets characterized by advantageous selection. The matching markets we study can also be interpreted as being characterized by advantageous selection in the sense that buyers and sellers with a high willingness to pay for the brokerage service are assumed to be the ones that bring most value to the market through the Nash bargain. The crucial difference with insurance or credit markets, however, is that consumer selection does not directly affect the cost function of the service providers, but instead manifests through an externality across the market: buyers care which sellers are selected into the intermediary market and vice versa. This, in turn, affects the revenue function of the intermediaries.

Finally, there is vast body of research that investigates the inefficiencies in the US real estate brokerage industry that can be attributed to a lack of price competition among brokers. Seminal theoretical contributions that point out conditions under which fixed commission rates can be socially harmful are Yinger (1981), Crockett (1982) and Miceli (1992). Hsieh and Moretti (2003) provide supporting empirical evidence of significant social waste in the US brokerage industry due to excessive broker entry. Other recent contributions that structurally aim to quantify the entry distortions are Han and Hong (2011) and Barwick and Pathak (2015).[7] In the present paper, the entry distortion is evaluated when brokers do compete in pricing their services. It is particularly relevant to address this question today, given that the adoption of new information technologies seems to have intensified price competition among intermediaries, not only in real estate brokerage (e.g. USDOJ and FTC report 2007), but also for many other intermediate service providers, like travel agencies and stock brokers, as pointed out by Levitt and Syverson (2008a).

The remainder of this paper is organized as follows. Section 2 presents the theoretical model and results. Section 3 proposes the methodology to empirically implement the model. Section 4 describes the data and the in-

[7] In addition, there are several papers that investigate the question whether fixed commission rates could be the result of a competitive market outcome or are more likely to arise from (tacit) collusion among brokers, usually from a principle-agent perspective. Examples are Carroll (1989), Anglin and Arnott (1999), Yavas (2001), Miceli, Pancak and Sirmans (2007), Levitt and Syverson (2008a) and Fisher and Yavas (2010).

stitutional setting of the Belgian real estate brokerage industry. Section 5 presents the results for the model calibration and welfare counterfactuals. The final section concludes.

2 Model

Consider a four-stage static model of symmetric imperfect competition among brokers who offer a service of matching buyers and sellers in a market for a homogeneous good. The implications of allowing for heterogeneous product characteristics are discussed in the next section when the methodology to implement the model empirically is introduced. The timing of the model can be summarized as follows:

- Stage 1: N brokers (out of a potentially infinite amount) enter the market.
- Stage 2: Participating brokers simultaneously announce the fees charged to sellers and buyers in return for their service.
- Stage 3: N^S sellers and N^B buyers (out of a potential mass S enter the market through one of the brokers.
- Stage 4: M real estate transactions occur through the brokerage industry and the broker service fees are paid.

Assume that brokers, sellers and buyers are risk-neutral and that sellers and buyers have unit supply and demand, respectively. Sellers are heterogeneous in their reservation value of providing the good to the market through one of the brokers, denoted by s and assumed smoothly distributed by $F^S(.)$ with density $f^S(.)$ on $[s^L, s^H]$ with $s^H > s^L$. One can think of s as the reservation price of selling the good either through the brokerage market or the outside market (for-sale-by owner) net of the gain (or loss) due to decreased (increased) search costs when hiring a broker. So, sellers with a low value of s are the ones who gain relatively most from the brokerage service. Similarly, buyers are heterogeneous in their valuation of purchasing the good through the brokerage market, denoted by b and assumed smoothly distributed by $F^B(.)$ with density $f^B(.)$ on $[b^L, b^H]$ with $b^H > b^L$. One can again think of b as the reservation value for the good added by the gain (or loss) in search costs when purchasing through one of the brokers compared to searching for the good through the outside market. Buyers with a high value of b

are thus the ones that gain relatively most from the brokerage service. The outside option of not participating in the market through one of the brokers (that is, the expected payoff of participating through the outside market) is normalized to zero for both sellers and buyers.

Assume that the distributions of seller reservation values and buyer valuations for the good are public information. Individual seller and buyers types, however, are *ex ante* private information, when sellers and buyers decide upon market participation (stage 3), and they become revealed *ex post* once a buyer is matched to a seller (stage 4).[8] The remainder of this section recursively specifies the occurrence of events and reports the resulting outcomes for each stage of the model.

2.1 Individual transaction valuations (stage 4)

When sellers participate in the market by hiring a broker they are charged a fee that only has to be paid conditional on the good being sold by the hired broker. The fee possibly consists of a flat component T and a percentage fee t charged proportional to the sales price of the property, p. The individual transaction value of a seller type s can hence be written as:

$$(1-t)p - s - T \tag{IV.1}$$

where p denotes the transaction price. Buyers are not directly charged for the broker service and the individual transaction value of buyer type b can therefore be written as:

$$b - p \tag{IV.2}$$

The fee charged to the seller, however, can (partially) be passed through in the bargain over the sales price between the buyer and the seller. More

[8]Note that the good traded in the market is implicitly assumed to be homogeneous. In practice, goods traded in one-to-one matching markets such as housing units or jobs of course consist of many attributes for which buyers and sellers might have heterogeneous valuations. For now, one can think of the model as being applicable to a market for a single good (with possibly multiple characteristics) for which buyers and sellers have heterogeneous preferences. Issues concerning aggregation to a market with many multi-characteristic goods are further discussed in sections 4 and 5 when the model is applied to the case of real estate brokerage.

specifically, assume the transaction price is chosen to maximize an asymmetric Nash bargain:

$$\max_p (b-p)^{(1-\beta)}((1-t)p - s - T)^{\beta} \qquad \text{(IV.3)}$$

where $\beta \in [0,1]$ denotes the bargaining weight of sellers and $1-\beta$ is the bargaining weight of buyers.[9] This yields the following expression for the transaction price:

$$p(b,s) = \beta b + (1-\beta)\frac{T+s}{1-t} \qquad \text{(IV.4)}$$

Nash bargaining implies that the transaction price is match-specific and depends on the valuations of the buyer and the reservation price of the seller that are being matched. The homogeneous real estate good is therefore allowed to be sold at dispersed prices, rather than being determined by a competitive market clearing mechanism, which would imply a single market price. This is consistent with the arguments of Stigler (1961) that price dispersion is inherent to markets with imperfect information and costly search, of which matching markets are a primary example.[10][11]

2.2 Buyer and seller participation (stage 3)

Assume that the service offered by brokers is perceived as differentiated across buyers and sellers, for example, by different locations of the brokers. Service differentiation is restricted, however, by the assumption that in equilibrium a symmetric and representative set of buyers and sellers is attracted by each broker. Market supply of sellers is equal to $N^S = \Sigma_{i=1}^{N} n_i^S$, where n_i^S is the number of sellers attracted by broker i, which is assumed to be

[9] Note that in real estate markets, the broker, rather than the seller, usually bargains over the transaction price with potential buyers (or buyer-brokers). However, a seller-broker (buyer-broker) contract typically also explicitly specifies that the broker should represent the best interest of the seller (buyer) in this process, which is assumed to be the case here. More generally, this paper ignores any potential principle-agent problems concerning the seller-broker or buyer-broker relationship, as investigated, for example, by Rutherford, Springer and Yavas (2005) and Levitt and Syverson (2008b).

[10] See, for example, Baye, Morgan, and Scholten (2007) for a further discussion on the determinants of price dispersion in markets with imperfect information.

[11] It should also be noted that the specific assumptions that only sellers are charged for the brokerage service and that buyers are not directly charged does not drive any of the results. All the derived results are robust to only buyers being directly charged, or when both sellers and buyers are charged part of the fee, as formalized in appendix A.1.

the same across brokers: $n_1^S = .. = n_N^S = N^S/N \equiv n^S$. Similarly, market demand for buyers is equal to $N^B = \Sigma_{i=1}^N n_i^B$ where n_i^B is the number of buyers attracted by broker i, again assuming symmetry across brokers $n_1^B = .. = n_N^B = N^B/N \equiv n^B$.

Assume, in addition, that the matching technology offered by the brokers is efficient and random. That is, the number of matches established by every broker is equal to $\min[n^B, n^S]$, the match probability of sellers is $\min[n^B, n^S]/n^S \equiv m^S$ and the match probability of buyers is $\min[n^B, n^S]/n^B \equiv m^B$. It follows, by broker symmetry, that the equilibrium number of matches that occur through the brokerage market is equal to $M = \min[N^B, N^S]$. In what follows, broker subscripts i are omitted to minimize the notational burden.

Expected seller and buyer utility of participating through the brokerage market can be written as:

$$u^s = ((1-t)p(\bar{b}, s) - s - T)m^S = \beta((1-t)\bar{b} - s - T)m^S \qquad \text{(IV.5)}$$

$$u^b = (b - p(b, \bar{s}))m^B = (1-\beta)(b - \frac{T + \bar{s}}{1-t})m^B \qquad \text{(IV.6)}$$

where $\bar{b}$ denotes the expected buyer valuation for the good and $\bar{s}$ the expected seller reservation price, respectively:

$$\bar{b} = \frac{S}{N^B} \int_0^{N^B/S} (F^B)^{-1}(1-x)dx \qquad \text{(IV.7)}$$

$$\bar{s} = \frac{S}{N^S} \int_0^{N^S/S} (F^S)^{-1}(x)dx \qquad \text{(IV.8)}$$

Sellers participate when $u^S \geq 0 \leftrightarrow s \leq (1-t)\bar{b} - T \equiv \tilde{s}$, where $\tilde{s}$ denotes the reservation price of the marginal seller that participates through the brokerage market. Similarly, buyers participate when $u^B \geq 0 \leftrightarrow b \geq \frac{T+\bar{s}}{1-t} \equiv \tilde{b}$, where $\tilde{b}$ denotes the marginal buyer valuation. Market supply of sellers and buyers can thus be summarized as:

$$N^S = SF^S(\tilde{s}) = SF^S((1-t)\bar{b} - T) \qquad \text{(IV.9)}$$

$$N^B = S(1 - F^B(\tilde{b})) = S(1 - F^B(\frac{T + \bar{s}}{1-t})) \qquad \text{(IV.10)}$$

Expression (IV.9) shows that market supply of sellers depends negatively on the service fees T and t charged by brokers, as one would expect. In addition, seller supply depends positively on the expected buyer valuation $\bar{b}$. All else equal, when sellers expect that buyers with a higher valuation participate in the market, more sellers participate because they expect to receive a higher price for their property. This in turn implies that seller supply is characterized by a negative externality induced by the participation decision of buyers. As illustrated by expression (IV.7), the expected buyer valuation depends negatively on the number of buyers that participate. This is because the marginal buyer always has a lower valuation for the good than infra-marginal buyers. Hence, the participation of this marginal buyer drives down the average valuation of all the buyers that participate in the market. Expression (IV.10), similarly, shows that market demand for buyers depends negatively on the service fees. In addition, it is characterized by a negative externality induced by the participation decision of sellers, through the expected seller reservation price $\bar{s}$. Low reservation price sellers enter the market first and hence increased seller participation raises the sales price buyers expect to pay, which in turn reduces buyer demand.[12]

Using expressions (IV.9) and (IV.10) and the definitions of $\tilde{s}$ and $\tilde{b}$ allows us to write the market clearing fees T and t as functions of the marginal and average preference values of sellers and buyers:

[12]Note that there is another channel through which externalities can result from the participation decision of users on either side. As is clear from the expressions (IV.5) and (IV.6) for expected seller and buyer utilities, respectively, the match probabilities on both sides, $m^s = \min[N^B, N^S]/N^S$ and $m^b = \min[N^B, N^S]/N^B$, also depend on the participation decision of users on both sides. The assumption that the matching technology is efficient, however, will imply that profit-maximizing brokers always balance the market by attracting the same amount of buyers and sellers. This in turn implies that the match probabilities of both buyers and sellers are equal to 1 in equilibrium and that these additional externalities play no role. See, Goos, Van Cayseele and Willekens (2014) for a more general treatment of the implications of matching frictions on the optimal pricing behavior of platform businesses.

$$T = \frac{\tilde{b}\tilde{s} - \bar{b}\bar{s}}{\bar{b} - \tilde{b}} \tag{IV.11}$$

$$1 - t = \frac{\tilde{s} - \bar{s}}{\bar{b} - \tilde{b}} \tag{IV.12}$$

Expressions (IV.11) and (IV.12) can be interpreted as a system of inverse demand equations, in which $\tilde{b} = F^{B^{-1}}(1 - N^B/S)$ and $\tilde{s} = F^{S^{-1}}(N^S/S)$, as follows from (IV.9) and (IV.10), and $\bar{b}$ and $\bar{s}$ are given by expressions (IV.7) and (IV.8). In what follows, it is assumed that any equilibrium market allocation N^B, N^S is uniquely established through the two market clearing values of the pricing instruments T and t that follow from (IV.11) and (IV.12).[13][14]

2.3 Imperfect broker competition (stage 2)

Broker profits and welfare

Expected broker profits can be written as:

[13]In other words, it is assumed that brokers can always resolve the coordination problem they face to attract two distinct user groups in the presence of indirect network externalities. This coordination problem is well-known from the two-sided markets literature and various solutions were proposed, for example, by Caillaud and Jullien (2003), Weyl (2010) and White and Weyl (2015). We do not explicitly address the issue here, however, given that it is precisely an important part of the "job" of brokers in matching markets to resolve the coordination problem. Brokers can credibly commit to sellers to search for a buyer to their best effort, given that payments to the broker only occur when a transaction is actually established. By this logic, the coordination problem is less of an issue in markets where intermediaries are involved in the trading process between participants and charge conditional payments compared to the classic two-sided market examples where the platform has no direct control over the interactions between attracted user groups, like payment card networks or newspapers.

[14]Another important concern is that the specification of seller supply and buyer demand implicitly assumes that sellers and buyers cannot reject a match in the final stage of the model, not even when the ex post realized transaction valuation is negative, which might be the case for some high reservation price sellers and low valuation buyers. If they could reject, the expectations about trading partner types, b and s , as currently presented would not be correctly defined - they would have to be defined conditional on individual buyer and seller types. One way to address this issue in our static setting is to assume that there is an opportunity cost for buyers and sellers to participate in the market while remaining unmatched. Appendix A.2 derives supply and demand including such a cost and shows that the reduced form specification in the main text is consistent if this cost is sufficiently large such that the marginal buyer and seller are always willing to trade. This cost can be interpreted as the direct disutility buyers and sellers experience when participating in the market while failing to find a trading partner, or as a reduced form characterization of the discounted search costs participant have to incur to stay in the market for more than one period in a dynamic setting.

$$\pi = (AR - MC)\min[n^B, n^S] - FC \tag{IV.13}$$

where $MC \geq 0$ denotes a constant per-match cost incurred when matching buyers and sellers and $FC \geq 0$ denotes a fixed cost incurred to operate in the market by each broker, independent of the number of transactions that they help carrying out. AR is defined as the expected or average per-match revenue:

$$AR \equiv T + t\bar{p} \tag{IV.14}$$

in which $\bar{p} \equiv p(\bar{b}, \bar{s})$ denotes the expected transaction price of transactions that occur through the brokerage market, which by the symmetry assumptions is the same for all brokers. Using equations (IV.4), (IV.11) and (IV.12), the average transaction price can be written as:

$$\bar{p} = \beta\bar{b} + (1-\beta)\tilde{b} \tag{IV.15}$$

In addition, using expressions (IV.11), (IV.12) and (IV.15), the average per-match revenue can be written as a function of marginal and average user types on both sides of the market:

$$AR = \beta(\bar{b} - \tilde{s}) + (1-\beta)(\tilde{b} - \bar{s}) \tag{IV.16}$$

Given that the marginal and average buyer types ($\tilde{b}$ and $\bar{b}$, respectively) are strictly decreasing in the number of buyers attracted into the brokerage market and the marginal and average seller types ($\tilde{s}$ and $\bar{s}$, respectively) are strictly increasing in the number of sellers, expression (IV.16) implies that expected per-match revenue is strictly decreasing in both the number of buyers and sellers that participate in the market by hiring a broker.

To model symmetric imperfect competition among brokers in providing their service to the market, we follow the approach of, for example, Bresnahan (1989) and Weyl and Fabinger (2013) who capture the degree of imperfect competition by a single “conduct parameter”. To apply the approach to our setting, the assumption is made that strategic interactions among brokers are restricted such that average per-match revenue for individual brokers is strictly decreasing in the number of users attracted on both sides of the market, i.e. $dAR/dn^I < 0$ for $I = B, S$, which is the equivalent to assuming that firms face downward sloping individual demand curves. This implies that individual broker profits, given by expression (IV.13), are strictly decreasing in the number of users on one side of the market if the attracted number of users on that side exceeds the number of users attracted on the other side, i.e. $d\pi/dn^I < 0$ if $n^I > n^J$ for $I \neq J$. This in turn implies that any profit maximizing equilibrium must always be balanced, i.e. $n^S = n^B = n$ or, equivalently, $N^S = N^B = M$. If not, brokers can always raise profits by lowering the number of users on the long side of the market. This result directly follows from the assumption that the matching technology available to brokers is efficient and it conveniently allows us to convert the problem of brokers competing to attract users on two distinct sides into a problem where the brokers compete in a single quantity (n) by using one of the available pricing instruments (e.g. T). The other available pricing instrument (e.g. t) is simply adjusted to ensure the balanced market condition holds and therefore no longer needs to be considered as a strategic decision variable.

Following Weyl & Fabinger (2013), instead of explicitly modeling the interactions among competing brokers, we assume that in any imperfectly comptitive equilibrium the elasticity-adjusted Lerner index is set equal to a conduct parameter θ, which in the model satisfies:

$$\frac{AR - MC}{AR}\left(-\frac{dM}{dAR}\frac{AR}{M}\right) = \theta \tag{IV.17}$$

where $\theta \in [0, 1]$ when the broker services are substitutes, which is assumed to be the case. As formalized by Weyl and Fabinger (2013), this framework nests a broad range of imperfect competition models, among which monopoly or cartel ($\theta = 1$); Bertrand ($\theta = 0$); Cournot ($\theta = 1/N$); Bresnahan (1989)'s constant conjectural variations model ($\theta = (1 + R)/N$ where $dM/dn = 1 + R$); and symmetrically differentiated Nash-in-prices and monopolistic competition (for which θ is not a constant). In our setup, however, we do not derive explicit conditions for these models, given that none of the results hinge on the specific underlying model of imperfect competition. The only thing that matters here is that any outcome on the continuum between monopoly and Bertrand is a feasible imperfect competition equilibrium.

To evaluate market efficiency, the private market equilibrium is compared to the outcome determined by a Pigouvian planner that optimally chooses the number of sellers N^S and buyers N^B attracted in the brokerage industry to maximize total social value, taking the number of brokers that operate in the market as given. Total social value generated in the market is equal to the sum of total industry profits $\Pi \equiv \pi N$ and total consumer surplus CS, defined as the sum of total buyer and seller surplus, which can be written as:

$$CS = (\beta(\tilde{s} - \bar{s}) + (1 - \beta)(\bar{b} - \tilde{b}))\min[N^B, N^S] \tag{IV.18}$$

By combining expressions (IV.13), (IV.16) ,and (IV.18), total social value W simplifies to:

$$W = (\bar{b} - \bar{s} - MC)\min[N^B, N^S] - FC * N \tag{IV.19}$$

Given that $\bar{b} - \bar{s}$ is strictly decreasing in N^B and N^S, the Pigouvian planner always balances the market, i.e. $N^B = N^S = M$, because welfare is strictly decreasing in participation on the long side of the market. This again conveniently allows us to simplify the social optimization problem to a problem with a single decision variable, in this case M.

Private market outcome

Proposition 1 summarizes the private market equilibrium when an exogenous number N of symmetric real estate agents operate the market. This result follows from equation N^S and N^B to M in expression (IV.16) for average per-match revenue, differentiating with respect to M and substituting the solution in the imperfect competition equation (IV.17).

Proposition 1 *Optimal private broker behavior implies that the equilibrium number of matches M established through the brokerage market satisfies*

$$AR - MC = \theta(MS + ET) \tag{IV.20}$$

where MS denotes marginal consumer surplus, defined as dCS/dM, which can be written as:

$$MS = \beta \frac{F^S(\tilde{s})}{f^S(\tilde{s})} + (1-\beta)\frac{1 - F^B(\tilde{b})}{f^B(\tilde{b})} \tag{IV.21}$$

and ET refers to an "externality tax", raised to internalize the cross-side participation externalities in buyer demand and seller supply, which can be written as:

$$ET = \beta(\bar{b} - \tilde{b}) + (1-\beta)(\tilde{s} - \bar{s}) \tag{IV.22}$$

Expression (IV.20) shows that the mark-up of average per-match revenue over per-match cost is increasing in the conduct parameter θ, ranging from zero under Bertrand competition ($\theta = 0$) to $MS{+}ET$, which is the monopoly mark-up ($\theta = 1$). The first term, MS, denotes marginal consumer surplus, which in a standard monopoly model is equal to the inverse hazard rate (or semi-elasticity) of demand and coincides with the classic Cournot distortion. In the present setting, MS is equal to the weighted sum of inverse hazard rates of seller supply and buyer demand, where the weights are equal to the bargaining weight of users on these respective sides. This is intuitive: if one side possesses no bargaining power in determining sales prices, users on that side capture no surplus from transactions and hence no surplus can be extracted by brokers from that side, independent of the elasticity of demand or supply. The second term, ET, refers to an externality tax raised by brokers to internalize the negative cross-side externalities present in the market. That is, brokers want to avoid attracting too many buyers because more buyers imply a lower average buyer valuation for the good, which in turn is disliked by the sellers because they expect to receive a lower price for their properties. Similarly, too many sellers imply a high average reservation price of sellers, which is disliked by buyers because they expect to pay a higher price for the good. To account for this, brokers charge a higher mark-up than they would without externalities.

From expression (IV.22) it is clear that the magnitude of ET depends on the spread between average and marginal user types on both sides of the market or, in other words, on the degree of heterogeneity in user types. When buyers are, for example, homogeneous in their valuation, sellers are indifferent to which buyer they will be matched and the participation decision of the marginal buyer causes no externalities. In this case, $\bar{b} = \tilde{b}$ and the first term in ET disappears because there is no externality for brokers to internalize on the buyer side. In contrast, when dispersion in buyer valuations is large, the spread between the marginal and average buyer valuation will be large and the marginal buyer entails a large externality. The tax raised to internalize this externality is precisely the spread between the average and marginal buyer valuation, weighted by the bargaining strength of sellers. Similarly, the tax to internalize the externality on the seller side is equal to spread between the marginal and average seller (where the former has a higher reservation price than the latter which is disliked by buyers), weighted by the measure of buyer bargaining power.

To sum up, proposition 1 demonstrates that under Bertrand competition ($\theta = 0$) the mark-up of average per-match revenue over per-match cost is equal to zero, whereas under monopoly pricing ($\theta = 1$) it is equated to a weighted version of the classic Cournot distortion plus a tax imposed to internalize the negative cross-side externalities present in the market. Depending on the degree of competition among brokers in pricing their services, any mark-up between these two bounds is a feasible private market outcome. To evaluate the distortions that might arise from private broker behavior, we now turn to the socially optimal market outcome.

Socially optimal outcome

Proposition 2 summarizes the social optimum chosen by a Pigouvian planner. The result follows from equating N^S and N^B to M in expression (IV.19) for total social value and rewriting the first-order condition with respect to M. The socially optimal degree of broker competition is derived from equating the private and social first-order conditions.

Proposition 2 *At the first-best social optimum, the equilibrium amount of matches M^* established through the brokerage market satisfies:*

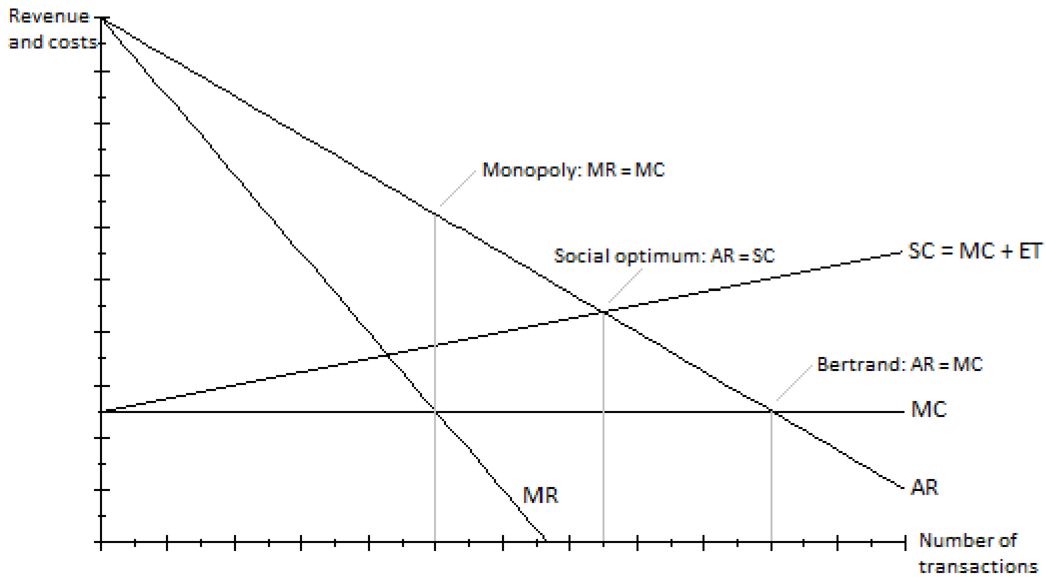
$$AR - MC = ET \tag{IV.23}$$

This implies that the socially optimal degree of competition among brokers in a private market satisfies:

$$\theta^* = \frac{ET}{MS + ET} \tag{IV.24}$$

Expression (IV.23) demonstrates that a Pigouvian planner also internalizes the selection effect by taxing the negative externalities induced by the participation decision of users on both sides. Furthermore, it does so exactly to the same extent a monopolist does in the private market. The externality tax is strictly positive in the presence of heterogeneity in buyer and/or seller types, which implies that Bertrand competition among brokers ($AR = MC$) is not socially optimal. In this case, broker fees are too low and the equilibrium number of matches is too high compared to the social optimum because the participation externalities present in the market are not properly internalized. The monopoly outcome, on the other hand, can never be efficient because, on top of the externality tax, broker fees are marked up by the weighted Cournot distortion, which results in upward distorted broker fees and hence insufficient participation of buyers and sellers. In a private market there thus exists an intermediate degree of imperfect competition θ^*, which establishes the first-best social optimum. Expression (IV.24) shows that θ^* depends on the magnitude of MS relative to ET. When marginal consumer surplus (the Cournot distortion) is small relative to the externality tax, the desired degree of market power is large and vice versa. Which of both measures prevails depends on the underlying distributions of user types and relative bargaining weights, as is clear from expressions (IV.21) and (IV.22).

Figure 1: *Graphical representation propositions 1 and 2*



Note: figure 1 assumes that market size is equal to one ($S = 1$), bargaining weights are symmetric ($\beta = 0.5$), seller reservation prices and buyer valuations are uniformly distributed on a unit interval ($s \approx U[0,1]$, $b \approx U[1,2]$) and $MC = 0.5$.

To further illustrate the intuition of propositions 1 and 2, Figure 1 graphically summarizes the results for linear buyer demand and seller supply. The number of transactions that occur through the brokerage market (M) are on the horizontal axis and broker revenues and costs are on the vertical axis. The AR curve illustrates that the expected per-match revenue of brokers decreases in the number of transactions that occur in the brokerage market. The marginal revenue curve, given by $MR = AR - MS - ET$, always lies below the average revenue curve. The Bertrand equilibrium is characterized by the point where the AR curve crosses the constant marginal cost curve and the monopoly (or cartel) equilibrium by the point where the marginal revenue curve crosses the marginal cost curve. As formalized in proposition 1, depending on brokers' market power measured by the conduct parameter θ, the private market equilibrium lies somewhere on the continuum in between the monopoly and Bertrand outcome. The social optimum is established at the point where the average revenue curve crosses the upward sloping social cost curve. The social cost of attracting buyers and sellers is equal to the marginal cost plus the tax raised to internalize the participation externalities of buyers and sellers: $SC = MC + ET$. In the presence of heterogeneity in buyer and seller types, the social optimum on the average revenue curve always lies in between the Bertrand and monopoly outcomes. So, there exists an intermediate degree of broker competition θ^* for which the incentives of the social planner and the private brokers are aligned, as formalized in proposition 2.

2.4 Free broker entry (stage 1)

Free entry equilibrium

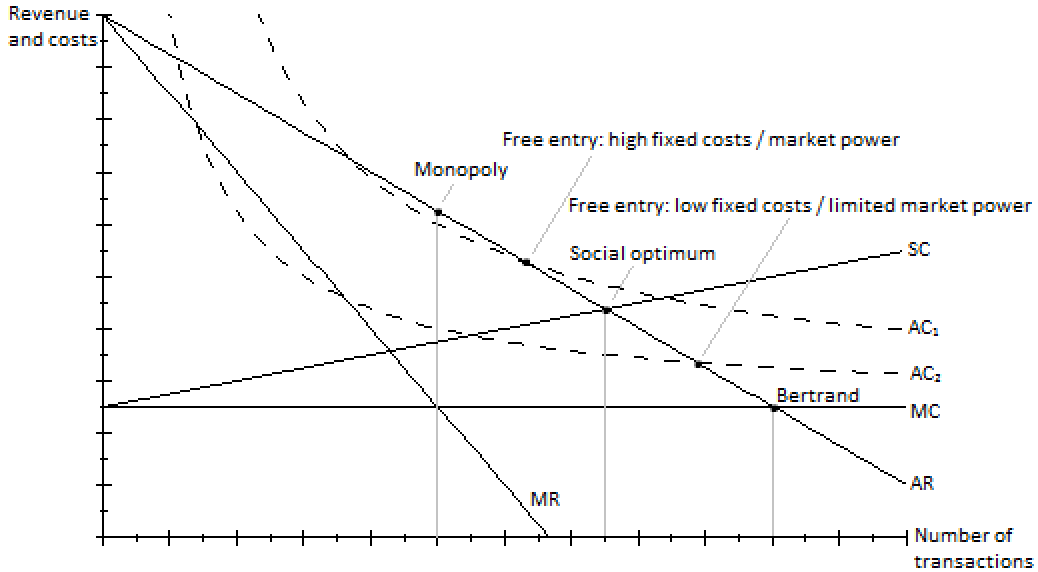
In the first stage of the model brokers can freely enter the market and they will do so as long as profits of the marginal entrant are weakly positive. Ignoring the integer constraint on the number of brokers, this implies that in a free entry equilibrium individual broker profits must be equal to zero:

$$\pi = (AR - MC)\frac{M}{N} - FC = 0 \Rightarrow N^e = \frac{(AR - MC)M}{FC} \tag{IV.25}$$

The number of brokers that enter the market depends on the mark-up they expect to receive in the second stage, given by expression (IV.20). When brokers, for example, collude on charging the monopoly service fee, expected per-match revenue (AR) and hence the number of transactions that occur through the brokerage industry (M) are independent of the number of brokers that enter the market. In this case, N is equal to $(AR - MC)M/FC$. Equation (IV.25), on the other hand, shows that Bertrand equilibrium ($AR = MC$) is not feasible in the presence of a positive fixed cost. More generally, when market power in the second stage is sufficiently large to cover the fixed cost of at least one entrant, the number of brokers that operate the market follows from the zero-profit condition (IV.25), where AR and M depend on N through the private first-order condition (IV.20). In what follows, the free entry equilibrium number of brokers that operate the market is denoted as N^{FE} and the average per-match revenue earned by brokers is always equal to the average per-match cost, $AR = AC$, where $AC = MC + FC/(M/N)$, as follows from rewriting expression (IV.25).[15] Greater market power in the second stage induces more brokers to enter the market, such that the number of transactions per broker (M/N) falls and hence the average cost incurred by each broker increases.

[15] Following Mankiw and Whinston (1986), the free entry equilibrium is unique when assumptions (a), (b) and (c) specified in proposition 3 below are satisfied.

Figure 2: *Graphical representation free entry equilibrium*



Note: in addition to the assumptions in figure 1, figure 2 assumes $FC = 0.01$ and that $N = 35$ for AC_1 and $N = 15$ for AC_2.

Figure 2 graphically illustrates the free entry equilibrium, which is characterized by the crossing of the average revenue curve and the average cost curve. Two cases are drawn. The average cost curve AC_1 crosses the average revenue curve above the social optimum, which implies that the average service fee is too high and too few transactions occur through the brokerage market compared to what is socially optimal. This case is more likely to occur when either fixed costs are high or when broker entry is high because brokers possess market power in setting their service fees (in stage 2), or both. In the extreme case where fixed operating costs are such that only one broker can enter the market, it will set the monopoly service fee and the AC curve will cross the AR curve at the monopoly outcome. Alternatively, when brokers collude to charge the monopoly service fee, the AC curve will also cross the AR curve at the monopoly outcome, even when fixed costs are relatively small. Many brokers will enter the market, which pushes up the AC curve, and every broker will only carry out a few but highly profitable transactions. In the second case, the average cost curve AC_2 crosses the average revenue curve below the social optimum, the average service fee is too low and too many transactions occur through the brokerage market because the negative participation externalities are not properly internalized. This is more likely to occur when fixed costs are small or when brokers possess limited market power in setting their service fees, or both.

Socially optimal entry

To evaluate how the private entry decision of brokers potentially distort the market outcomes, we follow Mankiw and Whinston (1986) by comparing the private free entry equilibrium to that of a social planner who optimally chooses the number of brokers that operate the market, taking private broker behavior once they enter the market as given. That is, the planner maximizes W, given by expression (IV.19) in which $N^B = N^S = nN$, by optimally choosing N, taking into account that the number of buyers and sellers attracted by individual brokers n is affected by N through the private first-order condition in the second stage of the model. The results are summarized in proposition 3.[16]

> **Proposition 3** *If for any N: (a) $dM/dN = n + Ndn/dN > 0$, (b) $Ndn/dN < 0$ and (c) $AR - MC > 0$, then the free entry equilibrium number of brokers N^{FE} strictly exceeds the socially optimal number of brokers, denoted by N^{SE}.*

The result that the private free entry equilibrium is always characterized by excessive entry is consistent with the findings of Mankiw and Whinston (1986), who demonstrate that, under the same set of assumptions (a)-(c), in standard oligopoly models there is always excessive entry in the presence of fixed costs. The intuition is that private brokers do not account for the fact that they "steal business" from the incumbent brokers. That is, when a new broker enters, the market expands (assumption (a)) in the sense that more matches will be established through the brokerage market, but if the market expansion is smaller than the individual number of matches established by the incumbent brokers prior to the entry decision of the marginal entrant, this entrant also steals business from the incumbent brokers (assumption (b)). Absent of fixed costs, business-stealing has no social cost, i.e. generated revenues in the market are simply divided among more brokers. In the presence of fixed costs, however, business-stealing implies that investments in fixed costs are wasted from a social point of view, given that the same market outcome could also be established by less brokers and hence less investments in fixed costs. The presence of fixed costs also implies that assumption (c) required for the result in proposition 3 to hold - that brokers

[16] The proof of proposition 3 can be found in appendix A.3.

charge a strictly positive mark-up over marginal cost - is satisfied.

2.5 Policy implications

The policy implications that follow from the results in propositions 1-3 are summarized in corollary 1. The first implication directly follows from combining the results in propositions 2 and 3. The second implication follows from the proof of proposition 3. The third implication follows from maximizing total social value in expression (IV.19) with respect to $M = N^B = N^S$, while also allowing for the number of brokers that operate in the market to depend on M through the free entry condition (IV.25), which is equivalent to maximizing consumer surplus, given by expression (IV.18).

Corollary 1 *(i) The first-best social optimum can be established by setting the service fees charged by brokers such that the average per-match revenue equates the social cost to attract buyers and sellers and by minimizing the number of brokers that operate in the market:*

$$AR = MC + ET \text{ and } N \to 0 \tag{IV.26}$$

(ii) When a social planner chooses the optimal number of brokers that operate in the market, while allowing them to privately compete in pricing their services once they have entered the market, the equilibrium number of matches M^{SE} established through the brokerage market satisfies:

$$AR = MC + ET + \frac{FC}{dM/dN} \tag{IV.27}$$

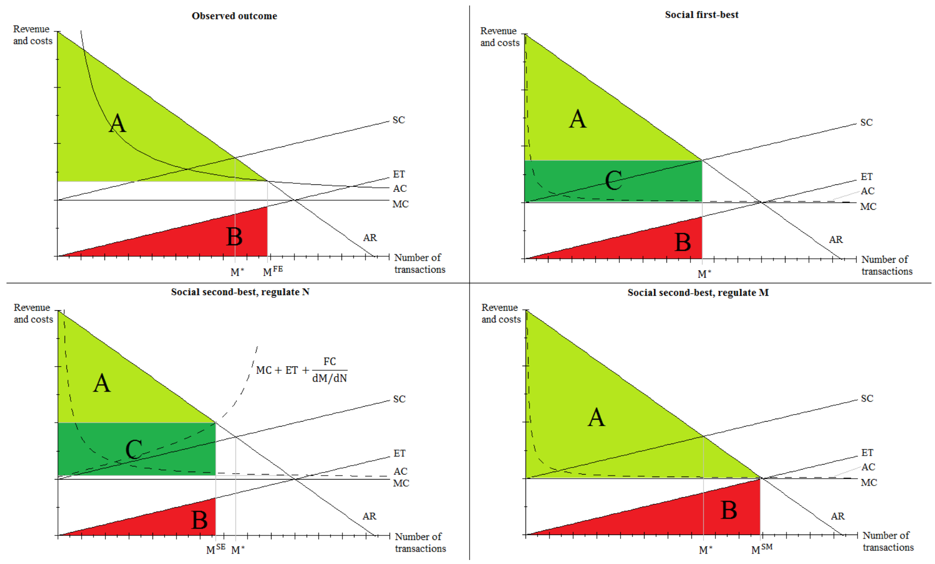
where the market expansion effect of the marginal entrant (dM/dN) follows from differentiating the private first-order condition (IV.20).

(iii) When a social planner sets the service fees to optimize the number of matches established in the brokerage market, while allowing brokers to freely enter the market, the equilibrium number of matches M^{SM} satisfies:

$$AR \to MC \text{ such that } N \to 0 \tag{IV.28}$$

Corrolary 1 shows the model outcomes when a social planner optimally regulates the service fees charged by brokers, broker entry or both. Figure 3 illustrates the welfare effects.

Figure 3: Policy implications



Note: in addition to the assumptions in figure 1, figure 3 assumes for the observed outcome that $FC = 0.01$ and $N = 15$, which implies that broker market power is $\theta = 0.13$. In the social first-best: $AR = SC$ and $N = 1$. In the social second-best when the social planner regulates entry: $AR = MC + ET + FC/(dM/dN)$ where dM/dN follows from the private FOC, imposing Bresnahan (1989)'s constant conjectural variations model $\theta = (1 + R)/N$. In the social second-best when the social planner regulates the number of transactions in the brokerage market: $AR = AC$ and $N = 1$.

As a benchmark, the top left panel of Figure 3 plots a possible observed free entry equilibrium. In this case, the number of matches M^{FE} is determined by the point where the average cost curve (AC) crosses the average revenue curve (AR). Social value generated in the brokerage market is equal to surface A below the AR curve minus surface B below the ET curve, where the latter captures the social cost of the externalities present in the market. In the free entry equilibrium brokers earn zero profits, which implies that all surplus generated in the market is consumer surplus divided among to buyers and sellers. The remaining three panels in figure 3 illustrate the implications of imposing the different policies described in corollary 1.

Firstly, when a social planner can regulate both brokerage service fees and market entry of brokers, implication (i) in corollary 1 applies. The planner equates average per-match revenue earned by brokers to the social cost of attracting buyers and sellers and minimizes the number of brokers to carry out the transactions. In the model, no integer constraint is imposed on the number of brokers and there are no constraints on the number of transactions a single broker can realize, so the socially optimal number of brokers approaches zero. Of course, in practice brokers have time constraints and there is a limit to the number of transactions a single broker can establish in a given time period. The planner should thus approximate the number of brokers required to realize the desired number of transactions, while minimizing the amount of business brokers steal from one another when operating in the market. The top right panel in figure 3 illustrates the social first-best when a single broker can realize all desired transactions. Social surplus generated by the brokerage industry is equal to consumer surplus (surface A minus surface B) plus the profits earned by the brokerage industry (surface C). Total social value is unambiguously higher compared to the free entry equilibrium, although there might be shift in surplus from buyers and sellers to the brokers when the average service fee in the social first-best is higher than in the free entry equilibrium - as it is drawn in figure 3.

Secondly, when a social planner can only influence the entry process of brokers, but not their pricing behavior once they have entered the market, implication (*ii*) in corollary 1 applies. Expression (IV.27) shows that the mark-up earned by brokers in this case is higher compared to the social first-best. The social planner not only internalizes the externalities induced by the participation decision of buyers and sellers (ET), but also the fixed costs that brokers incur to operate in the market (FC) divided by the market expansion effect of the marginal entrant (dM/dN). The additional mark-up is larger when fixed entry costs are larger and when the market expansion effect ($dM/dN = n + Ndn/dN$) relative to the business-stealing effect (Ndn/dN) is smaller or, in other words, when the social cost induced by the marginal entrant is higher. Note that to implement this policy, the planner has to know how the optimal pricing behavior of brokers is affected by changes in the number of brokers that operate in the market, i.e. how θ is affected by N. The bottom left panel of figure 3 illustrates the outcome, imposing Bresnahan (1989)'s constant conjectural variations model: $\theta = (1+R)/N$, where R is calculated from the free entry equilibrium and is assumed to remain constant as the number of brokers changes. The figure demonstrates that consumer surplus is smaller and profits of the brokerage industry are larger compared to the social first-best.

Thirdly, implication (*iii*) in corollary 1 applies when the social planner can influence the pricing behavior of brokers, but not their entry decision. In this case, independent of the mark-up chosen by the planner, brokers enter the market until they all earn zero profits and hence brokers bring no surplus to the market. The planner therefore maximizes total consumer surplus, which is strictly increasing in M, as follows from expression (IV.18). So, it is optimal to set average per-match revenue arbitrarily close to the per-match cost ($AR \rightarrow MC$), which minimizes the number of brokers that enter the market ($N \rightarrow 0$). Again, in practice the social planner should account for the time constraints of brokers and should target the service fees such that a minimal number of brokers enter the market to realize the desired transactions. The bottom right panel of figure 3 illustrates the outcome when the service fees are set such that a single broker enters the market. The figure demonstrates there are no broker profits in this case and that more buyer and sellers participate in the market compared the social first-best. This comes at the expense, however, of a higher social cost due to the externalities present in the market (surface B) and therefore total surplus is smaller compared to the social first-best.

In general, and not just for the case drawn in Figure 3, interventions (*i*), (*ii*) and (*iii*) are always (weakly) welfare improving compared to any observed free entry equilibrium, which can be anywhere on the continuum between the monopoly and Bertrand outcome, as discussed above. The greatest welfare gain is always established when imposing the social first-best (case (*i*)). However, which of the second-best cases (*ii*) or (*iii*) generates the largest welfare gains is ambiguous and depends on the parameters of seller supply, buyer demand, and on the cost structure of brokers. It is therefore essentially an empirical question. The remainder of this paper empirically applies the model to the case of the real estate brokerage and further discusses the practical implications of the theoretical results.

3 Empirical methodology

This section presents a methodology to quantify the parameters of the theoretical model.[17] It is assumed that the following cross-sectional data are

[17]As a reminder, the exogenous parameters in the model are the parameters of the distributions of buyer demand and seller supply ($F^B(.)$ and $F^S(.)$, respectively), seller bargaining weight (β), market size (S), broker per-match (MC) and fixed (FC) costs and the parameter(s) of the underlying model of broker competition that determine broker market power (θ). The endogenous outcome variables are the number of transactions that occur in the brokerage market (M) and the number of brokers that operate in the market (N).

available for one or multiple local markets in which brokers compete for transactions (e.g. a city in the case of real estate brokerage) within a given time frame (e.g. one or multiple years). Firstly, at the market-level: the number of transactions carried out by the brokerage industry (M) relative to the potential number of transactions (S); the number of brokers that operate in the market (N); and some (in)direct measures of broker costs (MC and FC) - e.g. Hsieh and Moretti (2003) use the wage earned by employees in other service industries within local markets as a proxy for the opportunity cost to operate as a real estate broker, but direct cost measures are preferred. Secondly, at the transaction-level: a representative sample of brokered transactions, with details on the (average) service fees charged by the brokers; sales prices and product characteristics of the traded good; and some measures of buyer and seller characteristics. Finally, it is useful to observe some broker characteristics or to observe multiple transactions carried out by the same broker to control for broker heterogeneity, as they are assumed to be homogeneous in the model.

3.1 Parametric specification of seller supply and buyer demand

Assume that buyer valuations are uniformly distributed over the interval [b^L,b^H] and that seller reservation prices are uniformly distributed over the interval [s^L,s^H]. The model then implies that buyers with a valuation in the range [$\tilde{b}$,b^H] participate in the brokerage market, where $\tilde{b}$ is the valuation of the marginal buyer. Sellers, on the other hand, participate when their reservation price is in the range [s^L,$\tilde{s}$], where $\tilde{s}$ is the reservation price of the marginal seller. Given that market participants are assumed to be randomly assigned to one another, it follows that the prices at which the good is sold are distributed by a symmetric triangular distribution.[18] The lowest possible price at which the good is sold occurs when buyer type $\tilde{b}$ is matched to seller type s^L. The sales price is then equal to $p(\tilde{b}, s^L) = \beta\tilde{b} + (1-\beta)(T+s^L)/(1-t) \equiv p^{\text{MIN}}$, which is observed with probability zero. Similarly, the highest possible sales price is $p(b^H, \tilde{s}) = \beta b^H + (1-\beta)(T+\tilde{s})/(1-t) \equiv p^{\text{MAX}}$, again observed with probability zero. The average sales price is the average of the minimum and

[18] To see this, note from expression (IV.4) that the sales price is a weighted sum of the buyer valuation and the seller reservation price and it is a familiar statistical property that any weighted sum of two independent continuous uniform random variables is distributed by a symmetric triangular distribution (e.g. Grinstead and Snell 1997).

maximum price: $\bar{p} = (p^{\text{MIN}} + p^{\text{MAX}})/2$, which is most likely to be observed.

In addition, the market clearing flat fee T and proportional fee t satisfy expressions (IV.11) and (IV.12), in which the marginal and average buyer and seller valuations can be written as a function of the fraction of buyers and sellers that participate in the market and the distributional parameters of buyer and seller reservation values. Combining expressions (IV.11) and (IV.12) with those for the minimum and maximum sales prices therefore allows to solve for the four relevant distributional parameters b^L,b^H,s^L and s^H as a function of the market clearing service fees (T and t), the fraction of buyers and sellers that participate in the market ($M/S = N^B/S = N^S/S$), seller bargaining weight (β) and the minimum and maximum sales price (p^{MIN} and p^{MAX}). By the assumption of linear supply and demand this system of equations has an analytical solution.

The average flat and proportional service fee and the fraction of buyers and sellers that participate in the market are assumed to be observed. So, it remains to obtain a proxy for seller and buyer bargaining weights and the minimum and maximum price of properties sold in the brokerage market to derive the parameters of supply and demand. The main challenge to obtain a proxy for these measures using transaction data is that the theoretical model assumes that a homogeneous good is traded in the market, whereas in practice traded goods are often heterogeneous in many dimensions. In other words, the model assumes that all dispersion in sales prices, measured by the difference between p^{MAX} and p^{MIN}, can be attributed to heterogeneity in buyer and seller characteristics, while in practice a large part of dispersion in sales prices can also be attributed to differences in the characteristics of the good - e.g. the size, location and age of a real estate property. Therefore, we introduce a methodology that allows us to derive an upper and a lower bound on the dispersion of sales prices that can be attributed to heterogeneity in buyer and seller characteristics. To do so, we build on the hedonic pricing model of Rosen (1974) and the extension of Harding, Rosenthal and Sirmans (2003) that allows for bargaining among market participants. The estimated bounds on price dispersion can subsequently be used to obtain bounds on the range of feasible values for the parameters of supply and demand.[19]

[19] Note that we face a nonstandard identification problem. With data on prices and quantities of goods sold by firms in standard product markets there are many techniques

3.2 Estimating residual sales price dispersion

Consider the following imperfectly competitive hedonic pricing model:

$$p_{gsb} = X_g\alpha^G + X_s\alpha^S + X_b\alpha^B + e_g + e_s + e_b \tag{IV.29}$$

where p_{gsb} denotes the sales price of good g when being sold by seller s to buyer b. X_g denotes a vector of observable characteristics of the heterogeneous good sold in the market and α^G is the vector with corresponding coefficients that can be interpreted as the value a specific characteristic of the good contributes to the sales price of the good on average. In addition, as in Harding, Rosenthal and Sirmans (2003), and opposed to the competitive hedonic pricing model of Rosen (1974), it is assumed that not only the characteristics of the good influence the price at which it is sold, but also the characteristics of the buyer and seller involved in the transaction. The intuition is that not all values of the product characteristics (α^G) are always known to all buyers and sellers and these informational imperfections leave room for bargaining over the sales price. This is typically the case in markets that are thin because the traded good is very heterogeneous, as in real estate markets. The vector X_s contains seller characteristics and the corresponding coefficient vector α^S measures how much these characteristics contribute in determining the sales price. Similarly, the vector X_b with coefficients α^B captures how buyer characteristics contribute to the sales price. The residuals e_g, e_s and e_b capture unobserved heterogeneity in product, seller and buyer characteristics, respectively.

To estimate the upper bound on the dispersion of observed sales prices that can be attributed to buyer and seller characteristics, the following hedonic pricing regression can be estimated:

$$p_{gsb} = X_g\alpha^G + \epsilon_{gsb} \tag{IV.30}$$

where the dispersion in the error term ϵ_{gsb} is interpreted as the residual

available in the literature to estimate consumer demand and firm market power. See, for example, Bresnahan (1989), Perloff, Karp and Golan (2007) and Einav and Levin (2010) for reviews. These techniques do not account, however, for the role of intermediaries. In the empirical literature on two-sided markets, there are some papers that estimate market power of platforms in the presence of externalities among different types of consumer groups (e.g. Rysman 2004; Lee 2013; Jeziorski 2014), but they typically do not allow for bargaining among matched trading partners. Finally, Bajari and Benkard (2005) propose a more general methodology than ours to estimate parameters of consumer demand and seller supply using the hedonic approach, but without intermediaries.

dispersion in sales prices that can be attributed to buyer and seller characteristics and to unobserved heterogeneity in product characteristics. Thus, if all relevant characteristics of the good that influence the sales price would be observed, the variance of e_g in expression (IV.29) would be zero, and ϵ_{gsb} would solely capture buyer and seller heterogeneity. If not all relevant product characteristics are observed, the variance of e_g is positive, and ϵ_{gsb} overestimates the heterogeneity in sales prices that can be attributed to buyers and sellers. Therefore the dispersion of ϵ_{gsb} is interpreted as an upper bound for the dispersion in sales prices that comes from buyer and seller heterogeneity. The values of p^{MIN} and p^{MAX} that follow can be obtained by fitting the symmetric triangular distribution to the distribution of the residuals around the predicted value of the regression.

To estimate the lower bound we want to estimate how much the terms $X_s\alpha^S$ and $X_b\alpha^B$ in expression (IV.29) contribute to the variation in sales prices. If we would estimate equation (IV.29), however, the obtained coefficients for these terms would likely be biased in the presence of unobserved product heterogeneity. This because different buyers and sellers are expected to differently value product characteristic and hence the component e_g in the error term will be correlated with the regressors in X_s and X_b. More specifically, when $e_g = X_s\delta^S + X_b\delta^B + e'_g$, where δ^S and δ^B measure how much sellers and buyers value the unobserved product characteristics, e_g in expression (IV.29) is clearly correlated with X_s and X_b when δ^S and δ^B differ from zero. To solve this, we follow Harding, Rosenthal and Sirmans (2003) by introducing two symmetry assumptions. Firstly, that the valuation of identical buyers and sellers for the unobserved product characteristics is the same, that is $\delta^S = \delta^B$. Secondly, that the way identical buyers and sellers can influence the sales price through the bargaining process is the same in magnitude but opposite. That is, $\alpha^S = \alpha^B$, which implies that the amount by which a certain degree of education, for example, allows a seller to push up the sales price is the same as it allows a buyer with the same educational level to push it down. Accounting for this allows us to rewrite equation (IV.29) as follows:

$$p_{gsb} = X_g\alpha^G + \alpha(X_s - X_b) + \delta(X_s + X_b) + e'_g + e_s + e_b \qquad \text{(IV.31)}$$

in which the term $\alpha(X_s - X_b)$ estimates how much buyer and seller attributes contribute to the variation in sales prices through the bargaining process and the term $\delta(X_s + X_b)$ estimates the valuation of buyers and sellers for unobserved product characteristics. The values of p^{MIN} and p^{MAX} can now be obtained by fitting the symmetric triangular distribution to the predicted values of the term $\alpha(X_s - X_b)$ around the predicted value of the regression.

As a final step, consistent with the assumptions in the empirical specification, it is assumed that the Nash bargaining game in the theoretical model is symmetric. That is, the bargaining weight of both buyers and sellers is one half: $\beta = 1 - \beta = 0.5$. Using the estimated bounds for p^{MAX} and p^{MIN} then allows us to calculate bounds for the values of the distributional parameters of the model, as described in the previous subsection. In addition, if either MC or FC is observed, the other cost measure can be calculated by using the zero profit condition (IV.25). Then, using MC and the parameters of seller supply and buyer demand, broker market power (θ) can be calculated from the private first-order condition (IV.20), which closes the model.

4 Data

4.1 Transaction-level data

The main dataset used for the analysis is a sample of 26,986 residential real estate properties that were sold in Belgium through one of 97 real estate agencies of a large franchise system in the period 2005-2014.[20] Table 1 provides descriptive statistics on sales prices and service fees charged by brokers (the latter are only available since 2011).

Table 1: *Descriptive statistics sales prices and service fees*

Description	In model	Obs.	Mean	St. Dev.	P5	P95
Sales price (in €)	p	26,986	215,279	92,763	90,000	392,500
Flat service fee (in €)	T	9,367	2,786	1,844	0	6,050
Proportional service fee	t	9,503	0.03	0.01	0	0.042

[20] The sample is restricted to houses, excluding apartments, for which another 10,666 transactions are observed. The same qualitative results are obtained when only using apartments in the analysis below, or when using both houses and apartments. Using both complicates the regression analysis because some observable property characteristics might affect the price of houses and apartments differently. Therefore, we prefer to exclude apartments from the sample.

An average property is sold for €215,579, ranging from €90,000 at the 5th percentile to €392,500 at the 95th percentile. The average flat fee charged by brokers is €2,786, ranging from €0 to €6,050 and the average proportional fee is 3%, ranging from 0 to 4.2%. This implies that brokers charge on average a total service fee of €9,182 or a commission rate of 4.3% for an average priced property.[21]

In addition, the dataset contains the initial listing price when properties were first brought on the market, time-to-sell and a broad range of observable property characteristics, such as size, age and number of bedrooms. Importantly, the exact location of properties is also observed, which allows us to construct measures such as the distance to the closest city center, distance to the capital city (Brussels) or distance to the nearest train station. For about half of the transactions, the data also contains the previous address of the buyer of a property. Observing the previous location of residence of buyers and sellers allows us to construct indirect measures of buyer and seller characteristics using publicly available administrative data for local living areas, for example, on median income, age and educational level of the population.[22]

[21]Note that, especially compared to the US where brokers usually charge a fixed commission rate of 5 or 6% (e.g. Hsieh and Moretti 2003), the service fees in our sample show strong variation. Furthermore, the service fees are on average lower and brokers use various pricing strategies - for 8% of the transactions only a flat fee was charged, for 17% only a proportional fee and for 75% a combination of both. These observations suggest that price competition among brokers is stronger in the Belgian brokerage industry than in the US industry. We believe that the crucial institutional difference that makes the Belgian market more competitive than the US market is that buyers (almost) never hire a broker in their search for a real estate property and only sellers hire brokers to sell their properties. This allows brokers to supply their services more independently than in an MLS system where real estate agents rely heavily on their colleagues to sell properties and can be penalized when deviating from the conventional commission rate (e.g. Levitt and Syverson 2008a). The reason for the absence of buyer representation by brokers in the Belgian market is likely due to the fact that every real estate transaction has to be approved and concluded by a notary, who supervises that all legal and administrative requirements are satisfied. Notaries thus essentially take up the role of guiding buyers through the process of buying a property which is executed by private brokers in the US. Note that notaries also receive a fee for this service, however, they are not private entities. Both the number of notaries and the fee they can charge for their service is highly regulated.

[22]Table B.1 in the appendix reports the descriptive statistics for the observable property characteristics and table B.2 for the proxies of buyer and seller characteristics.

4.2 Market-level data

The transaction data are complemented with aggregate data on the Belgian real estate brokerage industry. More specifically, data were collected on total market size, market share of the brokerage industry, broker entry and broker costs. Table 2 summarizes the data using 2013 as the reference year.

Table 2: *Market-level data*

Description	In model	Mean	Range
Market size	S	123,652	
Market share brokers	M/S	0.56	
# Brokers	N	6,728	4,494 - 8,963
Marginal costs (in €)	MC	3,339	983 - 5,696
Fixed costs (in €)	FC	60,141	35,882 - 84,389

Firstly, as a measure of market size, the total number of real estate transactions that occurred in Belgium in 2013 is used, calculated from publicly available administrative data from Statistics Belgium. In total there were 123,652 registered real estate transactions, of which 80,491 were houses and 43,161 apartments. So, the measure of market size consist of all the properties that were sold through the brokerage industry plus all the properties that were sold in the outside market (for-sale-by-owner).[23]

Secondly, a proxy for the fraction of transactions that occurred through the brokerage industry is obtained from a survey conducted by the Policy Research Center for Housing. The survey questioned 10,000 households that were randomly selected from the civil register about their current housing status. In the period 2009-2013, 710 of these households purchased a real estate property and 397 of them, or approximately 56%, claim that they bought the property from a seller that was assisted by a real estate broker.

Thirdly, two measures for the number of brokers that operate in the Belgian real estate market were collected. The first measure comes from data provided by the professional association of real estate brokers in Belgium

[23]The measure of market side should be interpreted as a lower bound on the actual potential market size, given that the measure does not include properties that were put up for sale, but remained unsold. In addition, it is possible that there were some buyers and sellers that would have entered the real estate market under different conditions (e.g. should broker service fees have been lower), but eventually decided not to enter.

(BIV), which contains the address of all registered members on the 1st of January 2011, 2012 and 2013. The data show that the number of registered brokers remains stable over these three years and in 2013 there were 8,963 registered brokers. The advantage of this measure is that registration with the professional association is mandatory in Belgium, which implies that all persons who are licensed to broker real estate transactions are included. The problem, however, is that not all brokers who are included in the list are necessarily active (full-time), so this number should be interpreted as an upper bound. As a second measure for the number of brokers active in the market we collected data from the largest online real estate listing platform in Belgium (www.immoweb.be). On the 12th of December 2013, 3,303 real estate agencies had at least one active real estate listing on the website. Assuming that the number of real estate brokers per agency is similar to that in the BIV-data (approximately 1.36), there were 4,494 real estate brokers active on the listing platform. Of course, it is unlikely that every broker in Belgium had an active listing on that day, so this measure is interpreted as a lower bound on the number of brokers that operate in the market. The average of both measures is 6,728.

Finally, data were collected from various sources on the advertisement and administrative costs to sell a real estate property in Belgium, resulting in a proxy of €983 for the monetary per-match cost.[24] In addition, when assuming that brokers can freely enter the market and earn zero profits, the costs to operate as a broker should also include the income a broker could earn when practicing a different profession. As a proxy for this opportunity cost, the average yearly wage of employees working in other service sectors than real estate brokerage is used, which was equal to €48,525 in 2013, as reported by the National Bank of Belgium.

Using a measure for the per-match cost MC, the implied fixed cost FC can be calculated from the zero profit condition (IV.25). The key question, however, is whether the opportunity cost to operate as a broker should be included in the measure for marginal costs or for fixed costs. On the one hand, it can be argued that the opportunity cost reflects the value of time that brokers invest in selling real estate properties. In this case, the opportunity cost (divided by the average number of yearly transactions per broker) should be included in the measure for MC, which results in a proxy

[24] See table B.3 in the appendix for details.

of €5,696 for MC and of €35,882 for FC. The amount of €35,882 then serves as a proxy for the monetary operating costs that brokers incur on a yearly basis independent of the number of properties they sell. These might be costs linked to office space, office supplies, purchasing or leasing a car, obtaining the broker license, the franchise fee, etc. On the other hand, as is for example argued by Hsieh and Moretti (2003), when brokers can freely enter the market, they are likely to waste valuable time and other recourses while inefficiently competing for transactions, especially when broker commission rates are fixed, as is typically observed in the US. By this logic, in the extreme case when all broker time is unproductive, the opportunity cost should be fully included in the measure of fixed costs. When the estimated monetary per-match cost of €983 is used as a proxy for MC, the implied fixed cost FC is €84,389. This measure then includes both the opportunity cost and the other fixed monetary operating costs. It seems reasonable to assume, however, that at least part of the time spent by brokers is productive, especially in our setting where broker commission rates are not fixed. For the baseline calibration of the model, the average proxy for MC of €3,339 and for FC of €60,141 is used, which each include half of the opportunity cost.

5 Model calibration and welfare counterfactuals

Using the data described in the previous section, this section first calibrates the outcomes of the theoretical model by applying the empirical methodology proposed in section 3. Subsequently, different welfare counterfactuals are constructed and discussed. Finally, sensitivity analysis is provided using alternative measures for broker costs and the number of brokers who operate in the market.

5.1 Model calibration

The first step is to obtain values for the parameters of buyer demand (b^L, b^H) and seller supply (s^L, s^H). To do so, remember that an estimate is required for the dispersion in sales prices that can be attributed to buyer and seller characteristics. The proposed methodology in the previous section allows to estimate these bounds. For the upper bound, after estimating the hedonic pricing equation (IV.30), the top panel of Figure 4 plots a Kernel density of the residuals around the predicted value of the regression.[25] In addition, the figure plots the fitted symmetric triangular distribution that minimizes the distance between the kernel density and the fitted distribution. The implied minimum sales price (p^{MIN}) is €141,064 and the maximum sales price (p^{MAX}) is €289,494 around the average of €215,279. So, the spread in sales prices due to heterogeneity in buyer and seller reservation values is therefore estimated to be €148,430. For the lower bound, after estimating equation (IV.31), the bottom panel of Figure 4 plots a kernel density of the predicted values of the term $\alpha(X^S - X^B)$ around the predicted value of the regression and the corresponding fitted triangular distribution.[26] The estimate for the minimum price is €206,807 and for the maximum price €223,751, implying a spread of €16,944.

Table 3: *Estimated supply and demand parameters*

Parameter	Upper bound	Lower bound
b^L	61,547	197,729
b^H	326,601	227,986
s^L	98,081	193,766
s^H	355,259	223,124

The implied values for the parameters of supply and demand are reported in table 3. For the upper bound on sales price dispersion, the valuation of buyers ranges from €61,547 to €326,601 and the reservation price of sellers from €98,081 to €355,259. For the lower bound on sales price dispersion, the spread ranges from €197,729 to €227,986 for buyer valuations and from €193,766 to €223,124 for seller reservation prices. Note that by construction dispersion in buyer valuations and seller reservation prices for the upper bound is larger than for the lower bound, which implies that for the upper bound buyer demand and seller supply are relatively inelastic with respect to changes in the broker service fee compared to the lower bound. For the upper bound the average revenue curve, plotted in Figures 1-3 above, is therefore relatively inelastic and thus relatively steep. For the lower bound

[25] The results of estimating regression equation (IV.30) using OLS are reported in the first column of table C.1 in the appendix.

[26] The regression results of estimating equation (IV.31) using OLS are reported in the second column of table C.1 and table C.2 in the appendix.

the average revenue curve is flatter.[27]

Also note that the difference between the upper and the lower bound in the dispersion of sales prices attributable to buyer and seller heterogeneity is large, €148,430 versus €16,944, respectively. This suggests that in the estimation for the upper bound there are still many property characteristics that are unobserved. Similarly, for the lower bound, there are likely many other unobserved buyer and seller characteristics that influence the sales price of properties. For the upper bound the spread implies, for example, that if the average seller would be lucky and be matched with the highest valuation buyer, the property would sell for €252,386. If unlucky and being matched to the lowest valuation buyer that participates in the market, the property would only sell for €178,171. Similarly, for the lower bound the property of the average seller would sell for a price ranging between €219,515 and €211,043. Intuitively, the spread of about €8,500 that can be attributed to "luck" in meeting the best trading partner in the lower bound perhaps comes closer to reality than the spread of about €74,000 implied by the upper bound. In what follows, the results are always reported for both the upper and the lower bound.

As a second step, various outcome variables of the model can be calculated by combining the obtained values for the parameters of supply and demand with the market-level data reported in table 2. Table 4 shows the calibrated values for the outcome variables that determine the optimal private service fee, as described in proposition 1.

[27] More specifically, the estimates for the parameters of supply and demand imply that the elasticity of the average revenue curve at the observed outcome is 0.04 for the upper bound on price dispersion and 0.37 for the lower bound. This implies that an increase in the average commission rate from the current 4.3% to 5.3%, for example, decreases the number of transactions in the brokerage market by 1% for the inelastic and by 9% for the elastic AR curve, respectively.

Table 4: *Baseline calibration*

Variable	Inelastic S & D	Elastic S & D
AR (in €)	9,182	9,182
MC (in €)	3,339	3,339
MS (in €)	146,224	16,692
ET (in €)	73,112	8,346
θ	0.026	0.233
W (in €)	5,062,680,199	577,929,349

The table shows that the observed total service fee ($AR =$€9,182) is significantly above marginal cost ($MC =$€3,339). As shown by expression (IV.20) in proposition 1, in a private market this mark-up is comprised of the sum of marginal surplus MS and the tax raised to internalize the participation externalities of buyers and sellers ET, weighted by the measure of broker market power θ. For the estimated upper bound of sales price dispersion, which corresponds to inelastic seller supply and buyer demand, $MS =$€146,224 and $ET =$€73,112 are relatively large compared to when supply and demand are elastic, $MS =$€16,692 and $ET =$€8,346. Given that expression (IV.20) is assumed to hold as an identity, corresponding broker market power is relatively small for inelastic compared to elastic supply and demand ($\theta = 0.026$ and $\theta = 0.233$, respectively). Finally, total social value generated by the Belgian real estate brokerage industry in 2013 is estimated to be about 5 billion Euro for the inelastic and 578 million Euro for the elastic bound on supply and demand.

5.2 Welfare counterfactuals

In this subsection, the observed private market outcomes described in table 4 are compared to those determined by a social planner. The three scenarios described in corollary 1 are considered and the results are summarized in table 5.

Table 5: *Welfare counterfactuals*

Scenario	Variable	Inelastic S & D	Elastic S & D
Regulate service fee, $N = 1$	M/S	0.431	0.518
	AR	59,634	11,059
	$\bar{p}$	240,885	216,231
	$AR/\bar{p}$	0.237	0.051
	N	1	1
	W^*	6,002,970,151	988,995,085
	Π	3,001,455,005	494,467,472
	CS	3,001,515,146	494,527,613
	W^*/W	1.185	1.711
Regulate entry, flexible service fee	M/S	0.427	0.408
	AR	61,310	15,996
	$\bar{p}$	241,736	218,737
	$AR/\bar{p}$	0.248	0.073
	N	517	2.261
	W^*	5,971,346,023	808,151,940
	Π	3,029,111,538	501,952,042
	CS	2,942,234,485	306,199,898
	W^{SE}/W	1.179	1.398
Regulate service fee, flexible entry	M/S	0.575	0.691
	AR	3,340	3,340
	$\bar{p}$	212,314	212,313
	$AR/\bar{p}$	0.016	0.016
	N	1	1
	W^*	5,335,986,832	879,120,106
	Π	0	0
	CS	5,335,986,832	879,120,106
	W^{SM}/W	1.053	1.521

The top panel of table 5 corresponds with case (i) in corollary 1 and reports the model outcomes when the social planner chooses the optimal number of transactions in the brokerage market, while minimizing the number of brokers that operate in the market. The results show that for both measures of inelastic and elastic supply and demand the planner attracts less buyers and sellers (a fraction of 0.43 and 0.52, respectively) compared to the observed private market outcome (where a fraction of 0.56 of the transactions occur in the brokerage market). This implies that the observed average service fee is below the socially desired level and the participation externalities of buyers and sellers are insufficiently internalized. The current commission rate is on average 4.3% and the optimal counterfactual commission rates are 5.1% for elastic and 23.7% for inelastic supply and demand, respectively. Assuming that all the transactions can be realized by a single broker, imposing the social first-best would imply a welfare gain ranging from 19% for inelastic to 71% for elastic supply and demand. Of course, in practice more than one broker is required to realize the desired number transactions. So, when appointing a realistic number of brokers the welfare gain would be smaller, as these brokers have to incur fixed operating costs. In addition, note that imposing the first-best implies a loss in consumer surplus allocated to buyers and sellers and the net gain comes from increased broker profits.

The middle panel of table 5 corresponds with case (*ii*) in corollary 1 and reports the model outcomes when the social planner determines the number of brokers that operate in the market, while allowing them to compete in pricing their services once they have entered the market. To do this, an assumption has to be made on how broker market power is affected when the number of brokers in the market changes. More specifically, Bresnahan (1989)'s constant conjectural variations model is imposed for which $\theta = (1 + R)/N$ where $dM/dn = 1 + R$. The conjectural variations parameter R can be calculated from the estimates of θ for the observed market outcome, as reported in table 4. In this case, the social planner reduces broker entry from the current 6,728 to 2,261 for elastic and to 517 for inelastic supply and demand. The corresponding commission rates increase from the current 4.3% to 7.3% and 25.3% respectively. This regulation entails an estimated welfare gain between 18% for inelastic and 40% for elastic supply and demand compared to the observed market outcome. Note that this policy implies an even larger loss in consumer surplus and a comparable gain in broker profits compared to the social first-best.

The bottom panel of table 5 corresponds with case (*iii*) in corollary 1. The counterfactual is constructed should the service fee be set such that exactly one broker enters the market. That is, the service fee is equated to the average cost of a single broker, which implies a commission rate of 1.6%. In this case, a fraction of 0.58 of the buyers and sellers participate in the brokerage market for inelastic supply and demand and a fraction of 0.69 for elastic supply and demand. Given that the counterfactual service fee is now below the observed service fee, there is a gain in consumer surplus of 5% to 52%. Broker profits remain zero, as the free entry condition continues to apply under this scenario.

Overall, table 5 suggests that the effectiveness of regulating broker entry or broker service fees crucially depends on how sensitive participation of buyers and seller is to changes in the service fee. For the inelastic bound, regulating broker entry is more effective than regulating the service fees, while the reverse holds for the elastic bound. In addition, when regulating broker entry, there can be important redistributive effects that shift surplus from buyers and sellers to brokers, which a social planner might want to take under consideration. If a regulator can only regulate entry, but nevertheless is only concerned with consumer surplus and not with broker profits, one possible solution is to sell licenses to brokers - i.e. impose a lump sum tax to operate as a broker. At a right price, this can induce the optimal number of brokers to enter the market under the second scenario in table 5, while broker profits would remain zero. The revenues of this taxation could then be redistributed to buyers and sellers through other real estate market policies.

More generally, lump-sum and per-match taxes can be used to achieve the desired market outcomes. If a regulator can levy both a per-transaction tax and a lump-sum tax to operate as a real estate agent in the market, it can always achieve the first-best outcome by combining these two instruments efficiently. It is important, however, that the regulator takes into account the strategic interactions among real estate agents when implementing a tax scheme. In the case of a monopoly or cartel ($\theta = 1$), levying a lump-sum tax to operate as a broker will reduce the number of brokers in the market and thus the social waste due to excessive entry, but will have no effect on the service fees charged by real estate agents. In this case, levying a fixed tax per-match will also reduce the number of real estate agents that operate in the market, but will at the same time raise the service fees further above their socially efficient level.[28] The regulator could, however, provide a subsidy to real estate agents to lower the service fees charged to the socially efficient level and could compensate these costs by levying a lump-sum tax to operate as a broker to achieve the first-best outcome.

While raising a lump-sum tax to operate as a broker has no effect on the service fees charged in the case of a monopoly ($\theta = 1$), it may well have

[28]Recall that the service fees charged under the monopoly outcome are always above the socially efficient level.

an effect under other models of imperfect competition. In the case of Bresnahan (1989)'s constant conjectural variations model ($\theta = (1 + R)/N$), for example, raising a lump-sum tax will both decrease the number of real estate agents that operate in the market and raise the service fees charged by the remaining brokers. In this case raising a per-match tax will have similar effects. If a regulator can only implement one fiscal instrument, it has to assess the market situation carefully. Although the effects raising a lump-sum tax or a per-match tax are similar when the service fees charged by real estate agents are initially too low, this is not the case when the initial service fees charged are too high. Raising a lump-sum tax to operate as a broker will in this case reduce the social loss due to excessive entry, but will at the same time further raise the service fees further above the socially efficient fee. A per-match subsidy will in this case lower the service fees charged by brokers, but also results in more entry of real estate agents. These examples illustrate that a regulator has to carefully assess the market conditions when designing and implementing tax schemes.

5.3 Sensitivity analysis

As discussed in section 4, some of the parameter values used for the baseline calibration of the model might suffer from measurement error. This section discusses the sensitivity of the results with respect to deviations of the model parameters from their baseline values. To start, remember that the proxy for marginal costs includes half of the opportunity cost to operate as a broker, measured by the wage brokers could potentially earn when working in a different service sector. This implicitly assumes that half of the effort of brokers goes to productively selling real estate properties and half is unproductive effort spent on marketing their services and competing with other brokers for transactions. The middle two columns of table 6 present the model outcomes should all effort be unproductive. In this case the opportunity cost is fully included in the measure for fixed costs (€84,389) and only the monetary costs of marketing and selling a real estate property are included in the measure for marginal costs (€983). The final two columns present the opposite case where all effort is assumed to be productive ($FC =$€35,882 and $MC =$€5,696). The table shows that the results are robust to alternative specifications of broker costs. The observed service fee (€9,182) always remains too low compared to the socially optimal fee, although it comes very close to the social optimum for elastic supply and demand and the lower bound on marginal costs. Intuitively, the estimated welfare gains from all policy interventions are larger for the lower bound on marginal costs compared to the baseline case and they are smaller for the upper bound. The only qualitative difference compared to table 5 is that for the upper bound on marginal costs regulating broker entry is now more effective than regulating the service fee for both inelastic and elastic supply and demand.

Table 6: *Welfare counterfactuals - sensitivity with respect to MC*

		$MC = 983$		$MC = 5,696$	
scenario	variable	Inelastic S & D	Elastic S & D	Inelastic S & D	Elastic S & D
Regulate service fee, $N = 1$	M/S	0.435	0.557	0.426	0.478
	AR	57,867	9,292	61,402	12,827
	N	1	1	1	1
	W^*	6,129,218,905	1,145,638,920	5,877,982,929	843,805,245
	W^*/W	1.21	1.982	1.161	1.46
Regulate entry, flexible service fee	M/S	0.427	0.399	0.425	0.418
	AR	61,064	16,370	62,038	15,521
	N	701	2,556	316	1,783
	W^{SE}	6,067,995,065	837,666,228	5,866,594,606	766,480,880
	W^{SE}/W	1.198	1.449	1.158	1.326
Regulate service fee, flexible entry	M/S	0.58	0.743	0.568	0.637
	AR	984	984	5,697	5,697
	N	1	1	1	1
	W^{SM}	5,448,21,335	1,018,364,459	5,224,881,689	750,057,080
	W^{SM}/W	1.076	1.762	1.032	1.297

In addition, remember from table 2 that an upper and a lower bound on the number of brokers active in the market is observed and the average of both was used for the baseline calibration. Table 7 reports the model outcomes for the upper and the lower bound. Again, none of these alternative specifications qualitatively alter the conclusions of the baseline case. Finally, the robustness of the results was also tested for possible measurement errors in the parameter values of market size (S), brokerage industry market share (M/S) and buyer and seller bargaining weight (β). For reasonable deviations from their baseline values, none of these qualitatively alter the conclusions of the baseline specification and the results are therefore omitted.

Table 7: *Welfare counterfactuals - sensitivity with respect to N*

		$MC = 983$		$MC = 5,696$	
scenario	variable	Inelastic S & D	Elastic S & D	Inelastic S & D	Elastic S & D
Regulate service fee, $N = 1$	M/S	0.431	0.518	0.431	0.518
	AR	59,634	11,059	59,634	11,059
	N	1	1	1	1
	W^*	6,002,940,254	988,965,188	6,002,985,148	989,010,082
	W^*/W	1.185	1.711	1.185	1.711
Regulate entry, flexible service fee	M/S	0.426	0.407	0.426	0.407
	AR	61,352	15,997	61,357	15,996
	N	345	1,510	688	3,012
	W^{SE}	5,971,346,336	808,151,939	5,971,346,321	808,151,941
	W^{SE}/W	1.179	1.398	1.179	1.398
Regulate service fee, flexible entry	M/S	0.574	0.69	0.574	0.69
	AR	3340	3340	3340	3340
	N	1	1	1	1
	W^{SM}	5,355,966,901	879,100,175	5,335,996,830	879,130,104
	W^{SM}/W	1.053	1.521	1.053	1.521

6 Conclusion

This paper aimed to make two contributions. Firstly, to present a general model of imperfect competition among intermediaries that operate in one-to-one matching markets, in which the intermediaries are allowed to freely enter the market and flexibly compete in pricing their services. The model showed that some private broker market power is justified from a social perspective, such that the broker service fee properly internalizes the participation externalities of buyers and sellers. In addition, it showed that generally an excessive number of intermediaries operate in a private market compared to what is socially desirable. The second contribution is to derive policy implications from this setting and to quantify the effects of various counterfactual regulatory interventions using data from the Belgian real estate brokerage industry. The counterfactuals suggest that regulating broker entry is more effective when seller supply and buyer demand are relatively insensitive to changes in the service fee charged by brokers. In contrast, targeting broker service fees is more effective when supply and demand is elastic. A regulator should be cautious, however, about redistributive effects that shift surplus from buyers and sellers to brokers when regulating broker entry, whereas regulating service fees always increases consumer surplus.

Appendices

A Theory and proofs

A.1 Equivalence different fee structures

In the main text only sellers and not buyers are directly charged for the brokerage service. By the assumption of Nash bargaining, however, the service fee can partially be passed through to buyers. Because Nash bargaining is efficient, the model outcomes are independent to whether the service fee is charged to sellers or buyers. To see this, consider the opposite case than the one analyzed in the main text where only buyers and not sellers are directly charged. In this case, the individual transaction valuation of a buyer type b is equal to $b-(1+t)p-T$ and of a seller type s is equal to $p-s$. Nash bargaining implies that the transaction price when a buyer type b and a seller type s are matched is $p(b,s)=\beta(b-T)/(1+t)+(1-\beta)s$. The inverse demand equations can then be written as $T=(\tilde{b}\tilde{s}-\bar{b}\bar{s})/(\tilde{s}-\bar{s})$ and $1+t=(\bar{b}-\tilde{b})/(\tilde{s}-\bar{b})$ and the average sales as $\bar{p}=\beta\tilde{s}+(1-\beta)\bar{s}$. Combining these expressions yields the following expression for average revenue: $AR\equiv t\bar{p}+T=\beta(\bar{b}-\tilde{s})+(1-\beta)(\tilde{b}-\bar{s})$, which is identical to expression (IV.16) in the main text. So, expression (IV.13) for broker profits and expression (IV.19) for welfare are also identical and all optimal pricing results carry through independent to which side of the market the service fee is charged.

A.2 Supply and demand with opportunity costs

Denote the cost of participating in the market while remaining unmatched as z^S for sellers and z^B for buyers. Expected seller and buyer utility of participating through the brokerage market can then be written as:

$$\begin{aligned} u^S &= ((1-t)p(\bar{b},s)-s-T)m^S - z^S(1-m^S) \\ u^B &= (b-p(b,\bar{s}))m^B - z^B(1-m^B) \end{aligned} \tag{IV.32}$$

When the expected pay-off of participating in the outside market is normalized to zero, sellers participate in the brokerage market when $u^S\geq 0 \Leftrightarrow s\leq (1-t)\bar{b}-T-\frac{1}{\beta}\frac{1-m^S}{m^S}z^S\equiv\tilde{s}$, where $\tilde{s}$ denotes the reservation price of the marginal seller. Similarly, buyers participate when $u^B\geq 0 \Leftrightarrow$

$b \geq \frac{T+\bar{s}}{1-t} - \frac{1}{1-\beta}\frac{1-m^B}{m^B}z^B \equiv \tilde{b}$, where $\tilde{b}$ denotes the marginal buyer valuation. When all buyers and sellers accept their randomly assigned trading partner - for which conditions are derived below - the expected buyer valuation $\bar{b}$ and the expected seller reservation price $\bar{s}$ are defined by expressions (IV.7) and (IV.8) in the main text, respectively. Market supply of sellers is $N^S = SF^S(\tilde{s})$ and market demand for buyers $N^B = S(1 - F^B(\tilde{b}))$.

To save notation, denote $Z^S = \frac{1}{\beta}\frac{1-m^S}{m^S}z^S$ and $Z^B = \frac{1}{1-\beta}\frac{1-m^B}{m^B}z^B$, such that the system of inverse demand equations can be written as:

$$\begin{aligned} T &= \frac{\tilde{b}\tilde{s} - \bar{b}\bar{s} + \tilde{b}Z^S - tildesZ^B - Z^S Z^B}{\bar{b} - \tilde{b} + Z^B} \\ 1 - t &= \frac{\tilde{s} - \bar{s} + Z^S}{\bar{b} - tildeb + Z^B} \end{aligned} \tag{IV.33}$$

Using that $\bar{p} = \beta\bar{b} + (1-\beta)\frac{T+\tilde{s}}{1-t}$, broker expected per-match revenue, $AR \equiv t\bar{p} + T$, can be written as:

$$AR = \beta(\bar{b} - tildes) + (1-\beta)(\tilde{b} - \bar{s} - \frac{1-m^S}{m^S}z^S - \frac{1-m^B}{m^B}z^B \tag{IV.34}$$

in which $\tilde{b} = F^{B^{-1}}(1 - N^B/S)$ and $\tilde{s} = F^{S^{-1}}(N^S/S)$. Using that by broker symmetry $N^B = n^B N$ and $N^S = n^S N$ and that $m^S = \min(n^B, n^S)/n^S$ and $m^B = \min(n^B, n^S)/n^B$ it follows that average per-match revenue is strictly decreasing in the number of buyers and sellers attracted by each broker: $\delta AR/\delta n^B < 0$ and $\delta AR/\delta n^S < 0$. As in the main text, this implies that broker profits are strictly decreasing in the number of participants on the long side of the market and the private market outcome is always balanced: $N^B = N^S = M$. So, the expression for average per-match revenue simplifies to expression (IV.16) in the main text and all derived results are robust to the alternative specification of supply and demand with opportunity costs.

For all sellers and buyers to accept the match with any randomly assigned trading partner it suffices that the marginal seller accepts the match with the marginal buyer that participates and vice versa. For this, the following conditions need to be satisfied:

$$\begin{aligned} &\beta((1-t)\tilde{b}-\tilde{s}-T) \geq -z^S \forall N^S, N^B \in [0,S] \\ &(1-\beta)(\tilde{b}-\frac{T+\tilde{s}}{1-t}) \geq -z^B \forall N^S, N^B \in [0,S] \end{aligned} \tag{IV.35}$$

Using the expressions for T and $1-t$ and the balanced market conditions, these conditions simplify to:

$$\begin{aligned} \beta(\tilde{s}-\bar{s}) \leq z^S \forall M \in [0,S] \\ (1-\beta)(\bar{b}-tildeb) \leq z^B \forall M \in [0,S] \end{aligned} \tag{IV.36}$$

The spreads $(\tilde{s}-\bar{s})$ and $(\bar{b}-tildeb)$ are increasing in M, so the conditions must hold for $M=S$. The conditions imply that the opportunity costs z^S and z^B must be sufficiently large relative to the degree of heterogeneity in seller and buyer types, respectively. That is, when there is great heterogeneity in seller and buyer types, for the marginal seller with the highest possible reservation price to accept a match with the marginal buyer type with the lowest valuation and vice versa, the opportunity costs of refusing the match must be greater compared to when the degree of heterogeneity in buyer and seller types is small.

A.3 Proof proposition 3

Differentiating expression (IV.19) for total social value, in which $N^S = N^B = nN$, with respect to N yields:

$$dW/dN = (\tilde{b}-\tilde{s}-MC)(n+Ndn/dN) - FC \tag{IV.37}$$

Equating expression (IV.37) to zero, using the expressions for AR (IV.16) and ET (IV.22), that $n+Ndn/dN = dM/dN$ and rewriting yields expression (IV.27). Note that dM/dN can be written as a function of M by solving the private first-order condition (IV.20) for N as a function of M (the solution is unique by assumption (b)) before differentiating. So, expression (IV.27) can be written solely as a function of M (independent of N) and hence can be solved for the equilibrium number of matches M^{SE} at the social optimum.

The excessive entry result follows from adding to and subtracting from expression (IV.37) expression (IV.13) for individual broker profits, in which $n^S = n^B = n$ and $N^S = N^B = nN$, which after rewriting yields:

$$dW/dN = \pi + (AR - MC)Ndn/dN - ET(n + Ndn/dN) \qquad \text{(IV.38)}$$

Expression (IV.38) illustrates the distortions that result from free entry in the private market relative to the social optimum. Under free entry in the private market individual broker profits equate zero ($\pi = 0$), while entry is socially optimal when the impact of the marginal entrant on social welfare is zero ($dW/dN = 0$). So, expression (IV.38) implies that private entry and socially optimal entry coincide when the sum of the second and the third term equals zero. When the sum of these terms is negative, there is excessive entry. This because $d\pi/dN < 0$, so $dW/dN = 0$ only holds when the number of brokers is smaller than under private entry. By assumptions (a)-(c) and the fact that $ET > 0$, the last two terms in (IV.38) are strictly negative and hence the private free entry equilibrium is unambiguously characterized by excessive entry. QED

B Descriptive statistics

Table B.1: *Descriptive statistics variables used in the regression analysis*

Variable	Obs.	Mean.	St. Dev.
List price (in €)	22,339	239,204	100,374
Sales price (in €)	22,395	220,099	91,567
Days-on-market	22,270	113	112
Living surface (in sq. m.)	21,129	180	62.7
Lot size (in sq. m.)	21,952	695	855
Year of construction	22,097	1957	30.7
Terraced	22,395	0.34	0.47
Semi-detached	22,395	0.25	0.43
Detached	22,395	0.40	0.49
Terrace	20,209	0.71	0.45
Elevator	11,737	0.00	0.06
Central heating	22,315	0.71	0.45
Heating material: gas	22,319	0.60	0.48
Heating material: electricity	22,319	0.07	0.26
Condensing boiler	22,216	0.05	0.21
Underfloor heating	22,295	0.03	0.17
Glazing: single	22,337	0.38	0.48
Glazing: double	22,337	0.77	0.42
Glazing: triple	22,337	0.00	0.07
Kitchen: luxuriously finished	22,337	0.06	0.24
Kitchen: dishwasher	22,337	0.30	0.45
State: luxuriously finished	22,316	0.04	0.21
State: ready to move in	22,316	0.60	0.48
State: minor refreshments necessary	22,316	0.19	0.39
Various: fireplace in living	20,194	0.12	0.32
Various: alarm	20,194	0.05	0.22
Environment: residential	20,803	0.14	0.35
Environment: villa district	20,803	0.04	0.19
Dist. center village	22,395	0.98	0.88
Dist. Brussels	22,395	57.8	30.2
Travel time to Brussels (in minutes)	22,395	55.5	19.5
Dist. nearest city	22,395	12.9	7.72
Dist. highway	22,395	5.39	5.22
Dist. train station	22,395	3.82	3.67
Year of sale	22,394	2010	2.71

Note: The sample of brokered real estate transactions contains besides information on prices (sales and listing price) and liquidity (time-on-market) also a very detailed description of the features of every property. The characteristics reported do not only describe the size of every dwelling (terraced vs. (semi-)detached, interior space, lot size, # bedrooms, # garages, # bathrooms), but also provide detailed information concerning the heating system (type (central heating), material (gas, electricity,), elements (underfloor heating, accumulators,)), isolation (single vs. double vs. triple glazing), state of the dwelling (ready to move in, luxuriously finished,), and its environment (residential, villa districts,) and location (distance to different amenities, major cities). For several rooms, such as the kitchen (well-maintained, dishwasher, ceramic stove,), bathroom (bath, shower,) and basement (wine cellar,), the realtor registered the features present.

Table B.2: *Descriptive statistics buyer and seller characteristics*

	Buyers			Sellers		
Variable	Obs.	Mean.	St. Dev.	Obs.	Mean.	St. Dev.
Avg. age pop.	9,605	40.99	3.75	22,383	40.97	3.34
% Married	9,602	0.53	0.11	22,378	0.55	0.10
Avg. size household	9,602	2.34	0.31	22,378	2.40	0.25
Med. tax. inc.	9,597	22,053	3,596	22,346	22,908	3,544
% Higher education	9,616	0.28	0.08	22,393	0.28	0.08
% Owners	9,616	0.69	0.17	22,393	0.74	0.14

Note: To construct measures of buyer and seller characteristics administrative data are used at the level of statistical sectors in Belgium. The 19,781 statistical sectors in Belgium are the lowest territorial level at which Statistics Belgium gathers information and, on average, have a surface of 1.5km2, and are home to approximately 550 inhabitants and/or 240 households. Given that we know the exact location of every dwelling and the previous address for a subsample of buyers, we can spatially join the respective x- and y-coordinates with the appropriate statistical sectors using the spatial join module in Quantum GIS. From Statistics Belgium we retrieved yearly data on different demographic variables and taxable incomes for every statistical sector. We either observed or managed to calculate the average age of the population, the percentage of reference persons of households who are married, the average size of a household and the median taxable income for every local living area. From the Census 2011 we furthermore retrieved the percentage of the population that finished higher education (where higher education is defined as a university degree or higher) and the percentage of owner-occupied houses in the total housing stock.

Table B.3: *Marginal costs*

What?	Details	Costs (range)	Costs (mean)	Source
Listing	*www.immoweb.be*	€100-€150	€125	*www.immoweb.be*
Other promotional activities	"For sale" sign, advertisement in local newspapers	100	€100	Own estimate
Special information duty	Building permits,...	€20 - €100	€60	*www. okra.be*
Certificates	Energy Performance	€150-€450 (€200-€600, according to www.immoweb.be)	€350	
	Electricity	€120	€120	
	Soil	€50	€50	
	Oil fuel tank	€65-€225	€145	
Title of land		€25	€25	
Information from property/charge registers	Mortgage	€16.50	€16.50	*www.kadaster.be*
	Cadaster	€16.50	€16.50	
Total:			€983	

Note: Table B.3 provides an overview of the monetary costs incurred by real estate agents when selling a property. Information from www.immoweb.be, the largest online listing service in Belgium, suggests that a listing costs between €100 and €150. Other online listing platforms in Belgium are usually free of charge. We assume that other promotional activities, such as a "for sale" sign and advertisement in local newspapers and so on, cost another €100. Since real estate agents also help sellers gather the necessary documents these costs are also initially incurred by the real estate agent. From www.okra.be and www.immoweb.be we learned that these costs are in total between €430 and €1120. Information from property registers finally contribute another €33. Given all these costs we calculate a monetary per-match cost of €983 for a representative transaction.

C Regression results

Table C.1: *Regression results*

Variable	Inelastic S & D	Elastic S & D	Variable	Inelastic S & D	Elastic S & D
Semi-detached	7,479***	4,903***	State: ready to move in	14,842***	14,946***
	(790.6)	(1,183)		(771.9)	(1,152)
Detached	19,948***	17,423***	State: minor refreshments necessary	-2,267***	-3,917***
	(1,040)	(1,593)		(736.9)	(1,146)
Living surface	422.5***	363.2***	Various: fireplace in living	6,846***	5,992***
	(30.55)	(48.63)		(1,043)	(1,635)
Living surface sq.	-0.255***	-0.164	Various: alarm	20,341***	18,301***
	(0.0731)	(0.119)		(1,583)	(2,632)
Lot size	45.18***	48.57***	Environment: residential	15,673***	14,679***
	(1.368)	(2.078)		(1,068)	(1,703)
Lot size sq.	-0.00427***	-0.00472***	Environment: villa district	18,145***	13,731***
	(0.00026)	(0.00036)		(1,905)	(2,941)
Terrace	8,480***	6,797***	Dist. highway	-195.1	-655.1*
	(757.7)	(1,144)		(234.4)	(347.7)
Central heating	8,557***	7,635***	Dist. train station	-305.3	-208.5
	(1,271)	(1,874)		(250.4)	(385.6)
Condensing boiler	9,407***	10,747***	Dist. Brussels	-1,472***	-655
	(1,535)	(2,611)		(390.7)	(639.9)
Underfloor heating	20,322***	20,463***	Dist. Brussels sq.	6.689**	5.699
	(2,243)	(3,561)		(2.993)	(4.819)
Glazing: single	-9,799***	-10,387***	Observations	18,812	8,083
	(754)	(1,126)	R-squared	0.828	0.844
Glazing: double	707.3	1,493	# Bedrooms	YES	YES
	(832.8)	(1,236)	# Garages	YES	YES
Glazing: triple	10,687***	15,627***	Building period	YES	YES
	(3,801)	(5,471)	Other quality controls	YES	YES
Kitchen: luxuriously finished	10,450***	8,550***	Other location controls	YES	YES
	(1,564)	(2,299)	Municipality FE	YES	YES
Kitchen: dishwasher	11,318***	12,020***	Broker FE	YES	YES
	(775.1)	(1,207)	Year-District FE	YES	YES
State: luxuriously finished	27,548***	25,673***	Buyer-Seller characteristics	NO	YES
	(1,982)	(2,952)			

Note: Table C.1 presents the regression results. Whereas most hedonic house price analyses use a log-transformed dependent variable, house prices are not log-transformed here since the purpose is for the residuals to capture the price spread (in euros). For some independent variables, such as interior space and lot size, the regression is therefore augmented with a quadratic term to capture possible nonlinearities. The first column of table C.1 presents the estimated coefficients for the hedonic pricing regression (IV.30) without buyer and seller characteristics. Almost all coefficients show the expected signs and are statistically significant. For example, the sales price of a dwelling is positively related to its interior space and lot size, but an additional square meter is less valuable for a large dwelling than for a smaller one. Also observe that (semi-)detached houses are more expensive than terraces ones. Furthermore, note that all the characteristics that relate to the quality of the property show their expected signs.[29]

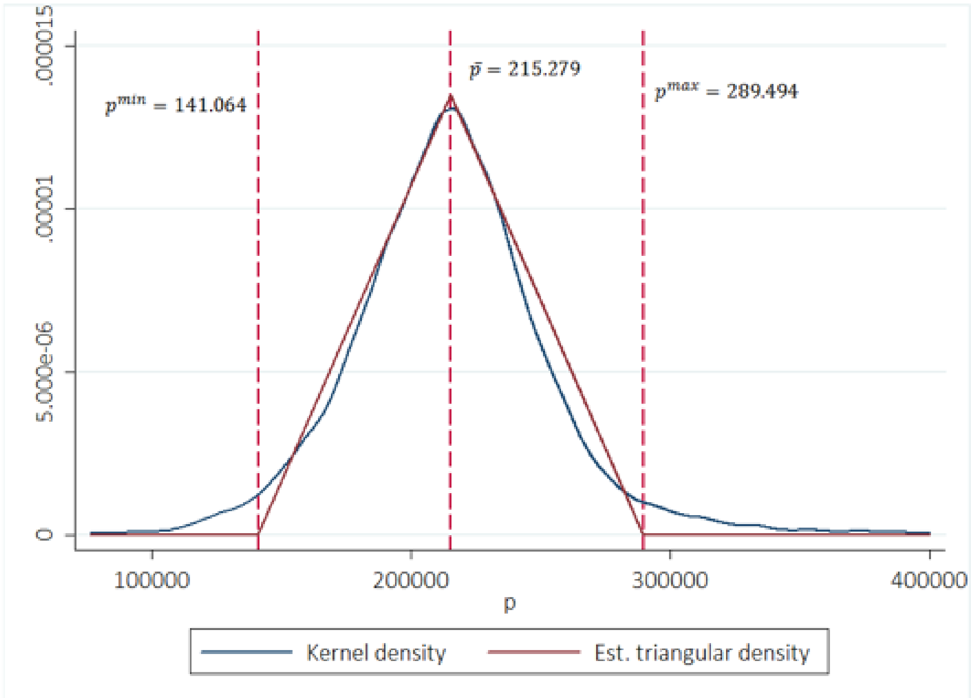
Following Harding *et al.* (2003), in the second column of table C.1 buyer and seller characteristics are included in the regression analysis. Harding *et al.* (2003), however, these variables are not observed at the individual level. Instead, their local living area counterparts are used as a proxy variable, as described in table B.2. In addition, table C.2 reports the regression coefficients for the "bargaining effect" (α) and the "demand effect" (δ) in expression (IV.31). There are significant positive demand effects from the percentage of people with a college education or higher, the percentage of people that is married, and the average age of the population. There is a significantly negative demand effects from the percentage of owners. The only significant bargaining effect comes from the percentage of people that enjoyed higher education, which suggests that the price of housing is increasing whenever the level of education of the seller is relatively high compared to that of the buyer.

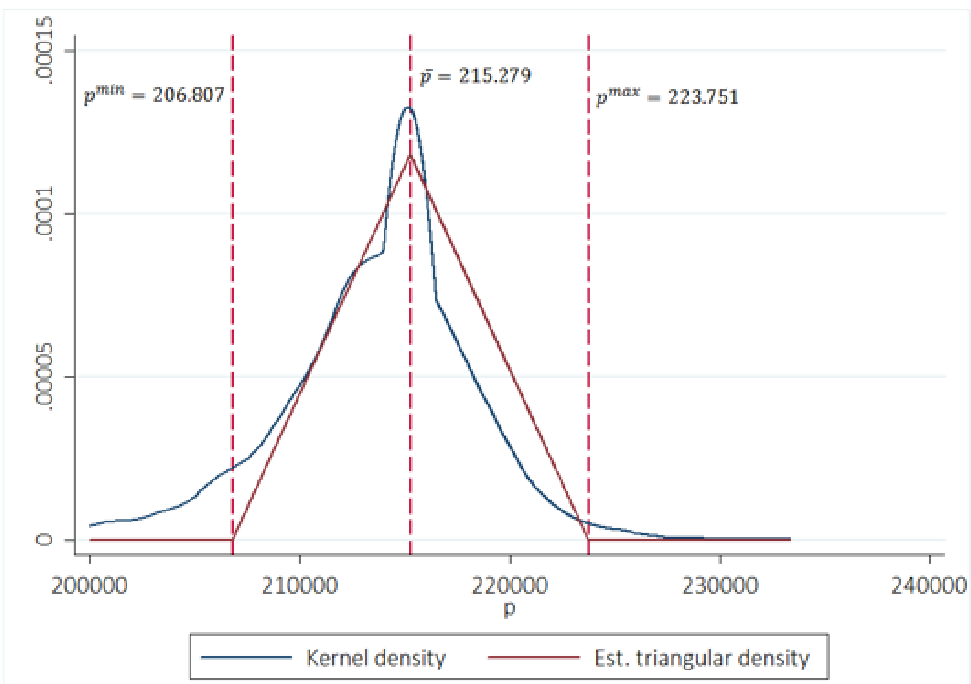
Table C.2: *Bargaining and demand effects*

Variable	Bargaining/ Demand	
Distance buyer-property	–	-53.79*
		(29.56)
% Higher education	α	47,868***
		(7,084)
	δ	63,127***
		(6,876)
% Owners	α	-6,002
		(4,688)
	δ	-14,544***
		(4,631)
% Married	α	5,664
		(11,155)
	δ	23,878**
		(10,410)
Avg. size household	α	1,121
		(3,946)
	δ	-5,762
		(3,700)
Avg. age population	α	252.3
		(162)
	δ	410.0***
		(155)
Ln(med. tax. inc.)	α	-5,422
		(4,456)
	δ	424.6
		(4,682)
Male	α	-17,144
		(11,658)
	δ	-12,111
		(10,412)

$*$ = 10%, $**$ = 5% and $***$ = 1% significance level

Figure C.1: *Estimated kernel and fitted triangular densities for Belgium as a whole*





Note: all procedures were carried out in Stata 11.2. For the lower bound we calculated the sum of squared deviations for all integer values ($p^{MAX} - \bar{p}$) between the minimum and the maximum of the estimated residuals. For the upper bound we performed the same procedure but used an interval of 10. For both the lower and the upper bound, the value of $p^{MAX} - \bar{p}$ that corresponds with the minimum value of the sum of squared deviations between the estimated Kernel density and the fitted triangular distribution is chosen.

General Conclusion

In this dissertation, I have presented four essays on local housing and real estate brokerage markets. As concluding remarks and summaries were already presented at the end of every chapter, this section contains some more general lessons and insights. Although this dissertation contains four chapters where we attempted to formulate interesting research questions and have tried to answer them using the appropriate data and techniques, combining them into a single thesis has also brought up some, which I believe to be, interesting new ideas for future research. These are loosely discussed in the next paragraphs.

In the introduction of this thesis we started with the observations of John Quigley and Duncan Maclennan. Already in 1979 the late housing economist John Quigley noted that the housing commodity is characterized by a peculiar combination of features that differentiate it from other goods and markets, such as the (US) equity market, where arbitrage is instantaneous and search costs are negligible. The inherent spatial fixity, durability and heterogeneity of the housing commodity has led Duncan Maclennan (1982) to conclude that "*the housing market is really not a single neoclassical exchange market, but is rather a set of overlapping submarkets differentiated by tenure, location, size and quality.*" Quigley (2002) also argued that these characteristics imply that transaction costs are substantial and significant in housing markets. These transaction costs do not only entail registration- and notary fees, but also search costs. Due to the heterogeneous nature of the housing good and spatial fixity, searching is costly both for buyers and sellers. It is in these types of environments that intermediaries, such as real estate agents, thrive. In the four essays presented in this dissertation we focus on the spatial dimension of the housing commodity and the inherent heterogeneity that characterize it.

In the third chapter of this dissertation we investigated the relationship between the housing stock composition at the neighborhood level and housing prices. While housing economists have unanimously acknowledged that housing is a heterogeneous good and methods, such as the hedonic pricing method (Rosen, 1974), have been developed to deal with this inherent heterogeneity, this chapter is one of (very) few studies investigating the effect of neighborhood housing stock composition on housing prices. The large heterogeneity in both dwellings and neighborhoods are explanations for the fact that housing markets are frequently characterized as "thin" markets. Especially in Belgium, where housing is mainly organized on a relatively small scale, i.e. individuals hire an architect and contractor to construct their preferred home, dwellings are extremely heterogeneous. This, in turn, implies that potential buyers/sellers have few reference points and it is not easy to observe a "market-clearing" price for many properties. This void is potentially filled by bargaining between buyers and sellers of properties. While the hedonic pricing model is extremely suitable for the analysis of housing prices, because of its capability to deal with the inherent heterogeneous nature of the housing good, its theoretical foundations do not allow for any form of bargaining between buyers and sellers. In 2003, Harding *et al.* (2003) published a paper in the *Review of Economics and Statistics*, where they develop (and apply) a novel identification strategy that can be used to uncover bargaining effects. While the method developed by Harding *et al.* (2003) was merely used as an instrument to quantify the structural parameters of buyer demand and seller supply in chapter 4, I believe that the unique nature of the housing commodity suggests a high demand for econometric models that allow for bargaining between agents.

As bargaining effects are especially likely to be observed in "thin" markets, it is perhaps also interesting to estimate a similar bargaining model for different markets and types of properties. In chapter 2, I already established that housing markets and - construction is organized very differently in the Netherlands than in Belgium. While housing construction is organized on a small scale (individuals hire an architect and contractor themselves) in Belgium and (spatial) planning regulations are loose, housing construction in the Netherlands is mostly organized on a large scale (a single housing construction project frequently entails hundreds of (highly) similar dwellings) and spatial planning is (more) strict. Obviously, these differences also result in a different housing stock. While neighboring dwellings in the Netherlands frequently look much alike, dwellings in Belgium can differ in many aspects from neighboring dwellings. This consequently implies that it might be more costly for a potential seller in Belgium to search for an appropriate (listing) price for their property, which leaves room for bargaining. Although the reasoning presented here is just a hunch, it can be empirically tested when the appropriate data are available.

While the previous two paragraphs have focused on aspects related to heterogeneity, I was also very much interested in the spatial dimension that is inherent to the housing commodity. Although I started out with an interest in econometric models that deal with spatial panel data, I personally experienced the increasing availability of spatial data over the last 4.5 years. Although one of the first applications of spatial analysis (in epidemiology) already dates back to the French geographer Charles Picquet (1832), who wrote the famous "*rapport sur la marche et les effets du choléra dans Paris et le département de la Seine*", the analysis of spatial data in (housing) economics has increased tremendously over the last two decades due to this increased availability of data and also computational power. In the third chapter of this dissertation we have shown that geospatial can be used to create additional variables that can be used in econometric models, such as the hedonic pricing models. I believe that the increased availability of spatial and other data, due to advances in ICT, is one of the most important developments for housing (economists) in recent decades. Although the so-called "Big Data" provides a whole new array of opportunities for researchers, it also presents some new challenges such as (lack) of computational power to analyze all the available data.

As different dwellings are obviously located in different locations with potentially different (neighborhood) characteristics, I welcome the initiatives of governments and institutions to provide more (spatial) data to researchers. As locations can also vary in a potentially infinite number of features, it is clear that there will always be unobserved (spatial) heterogeneity. Fortunately, econometricians have developed techniques and methods to deal with (un)observed (spatial) heterogeneity, by allowing for example for (spatially) correlated error terms (e.g. spatial error model (SEM)) or (spatial) dependent variables (e.g. spatial lag model (SAR)). Ever since the seminal contributions of Paelinck & Klaassen (1979) and Anselin (1988) spatial econometric models, such as spatial error- and spatial lag models, have transformed from an obscurity to a workhorse of modern (housing) economists.

While the theoretical foundations of spatial econometric models using cross-sectional data were already established in the eighties, this field of research remains highly active. In recent years, authors such as Pesaran (2004, 2007a, 2007b, 2011) and Baltagi (2011) have made important contributions in the field of econometrics with spatial panel data. In chapter 1 of this dissertation we showed that housing price spillovers between local markets follow distinct spatial patterns. The results presented furthermore suggested that the linguistic border in Belgium acts as a "hard" boundary in the peripherally located eastern and western parts of Belgium. While the first chapter contributes to the expanding literature in this particular field of research, we believe that these considerations can also be studied using micro-level data. In chapter 2 we already showed that higher housing prices in one market (the Netherlands) potentially spill over to neighboring markets in other countries. In a similar fashion it is also possible to study the "border effects" potentially arising from the linguistic border that splits Belgium into two distinct linguistic regions.

While the first three chapters of this dissertation focused on local housing markets, the last chapter investigates the market efficiency in the Belgian real estate brokerage market. Despite that this chapter is quite different from the other three chapters presented, it is important to emphasize that real estate markets and real estate brokerage markets are closely related. A simple observation that I believe proves this statement is the observation that housing prices and the number of (active) real estate agents in the market are frequently positively correlated. Real estate agents have traditionally acted as intermediaries between buyers and sellers in real estate markets and provide promotional services, search for eligible buyers and provide market information. The rise of modern information and communication technologies, such as the internet, in recent years will likely also affect (the role of) real estate agents. As the internet not only provides valuable market information, but also enables potential sellers to search for eligible buyers more easily themselves. All in all, the advent of modern ICT has lowered (monetary) search costs for both sellers and buyers of properties, which has decreased the information monopoly of real estate agents. It is no wonder that in recent years researchers (e.g., Bernheim & Meer, 2008; Hendel et al., 2009) have started investigating whether selling with the help of real estate agent yields "better" market outcomes, such as a higher sales price (after commission), probability of sale and/or a shorter time-on-the-market.

While there is, for now, little evidence that real estate agents achieve better market outcomes, it might be the case that certain real estate agents "structurally" achieve better market outcomes than others. In the last chapter we ignored the fact that the market outcomes achieved by real estate agents are potentially a function of the commission fees charged. As was rightfully acknowledged by Levitt & Syverson (2008b), the service fees charged by real estate agents serve as an incentive mechanism. Levitt & Syverson (2008b) showed that the observed fixed commission rates in the US, do not adequately incentivize real estate agents. Real estate agents tend to sell properties of their clients more quickly, and at a lower sales prices, than the properties they own themselves. This might also help explaining the results found by Hendel *et al.* (2009). As variation in commission rates in Belgium is observed, it is particularly interesting to investigate whether this variation can help explaining differences in market outcomes. Together with my highly appreciated co-worker Bert Willekens, I hope to be able to start digging deeper into these issues in the near future.

References

1. Abbott, A. and G. De Vita (2013). Testing for long-run convergence across regional house prices in the UK: a pairwise approach. *Applied Economics*, 45(10), 1227-1238.

2. Ahlfeldt, G.M. and N. Holman (2015). Distinctively Different: A New Approach to Valuing Architectural Amenities. CESifo Working Paper Series No. 5221.

3. Ahlfeldt, G.M. and W. Maennig (2010). Substitutability and Complementarity of Urban Amenities : External Effects of Built Heritage in Berlin. *Real Estate Economics*, 38(2), 285-323.

4. Ahlfeldt, G.M. and A. Mastro (2012). Valuing Iconic Design: Frank Lloyd Wright Architecture in Oak Park, Illinois. *Housing Studies*, 27(8), 1079-1099.

5. Albrecht, J., Gautier, P.A. and S. Vroman (2016). Directed search in the housing market. *Review of Economic Dynamics*, 19, 218-231.

6. Alexander, C. and M. Barrow (1994). Seasonality and cointegration of regional house prices in the UK. *Urban Studies*, 31(10), 1667-1689.

7. Anderson, J.E. and E. van Wincoop (2003). Gravity with Gravitas: A Solution to the Border Puzzle. *American Economic Review*, 93(1), 170-192.

8. Anenberg, E. and P. Bayer (2015). Endogenous Sources of Volatility in Housing Markets: The Joint Buyer-Seller Problem. NBER Working Paper No. 18980.

9. Anglin, P. and R.J. Arnott (1999). Are brokers' commission rates on home sales too high? A conceptual analysis. *Real Estate Economics*, 27(4), 719-749.

10. Anselin, L. (1988). *Spatial Econometrics: methods and models*. Dordrecht: Kluwer Academic Publishers, pp. 284.

11. Arnott, R.J. (2013). *Regional and Urban Economics Parts 1 & 2*. Abingdon: Routledge, pp. 576.

12. Arraiz, I., Drukker, D.M., Kelejian, H.H. and I.R. Prucha (2010). A spatial Cliff-Ord-type model with heteroskedastic innovations: Small and large sample results. *Journal of Regional Science*, 50(2), 592-614.

13. Armstrong, M. (2006). Competition in Two-Sided Markets. *Rand Journal of Economics*, 37(3), 668-691.

14. Ashworth, J. and S.C. Parker (1997). Modelling regional house prices in the UK. *Scottish Journal of Political Economy*, 44(3), 225-246.

15. Bajari, P. and L. Benkard (2003). Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach. *Journal of Political Economy*, 113(6), 1239-1276.

16. Ball, M. (2003). Markets and the structure of the housebuilding industry, an international perspective. *Urban Studies*, 40(5-6), 897-916.

17. Baltagi, B.H. (2008). *Econometric Analysis of Panel Data*, Hoboken (New Jersey): John Wiley and Sons, pp. 366.

18. Barwick, P.J. and P.A. Pathak (2015). The Costs of Freen Entry: An Empirical Study of Real Estate Agents in Greater Boston. *RAND Journal of Economics*, 46(1), 103-145.

19. Baye, M., Morgan, J. and P. Scholten (2007). Information, Search and Price Dispersion. In: *Handbook on Economics and Information Systems*, T. Hendershott (ed.), Elsevier.

20. Benjamin, J.D., Jud, G.D. and G.S. Sirmans (2000). What Do We Know About Real Estate Brokerage?. *Journal of Real Estate Research*, 20(1), 5-30.

21. Biglaiser, G. (1993). Middlemen as Experts. *RAND Journal of Economics*, 24(2), 212-223.

22. Breitung, J. and M.H. Pesaran (2008). Unit Roots and Cointegration in Panels. In: *The Econometrics of Panel Data: Fundamentals*

and Recent Developments in Theory and Practice, László Mátyás and Patrick Sevestre (Eds.), Springer-Verlag Berlin Heidelberg, 279-322.

23. Bresnahan, T.F. (1989). Empirical Studies of Industries with Market Power. In: *Handbook of Industrial Organization, Vol. 2*, Richard Schmalensee and R. Willig (Eds.), Amsterdam: North-Holland, 1011-1057.

24. Brunsdon, C., Fotheringham, A.S. and M.E. Charlton (1996). Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis*, 28(4), 281-298.

25. Burnell, J.D. (1988). Crime and racial composition in contiguous communities as negative externalities: Prejudiced households' evaluation of crime rates and segregation nearby reduces housing values and tax revenues. *American Journal of Economics and Sociology*, 47(2), 177-193.

26. Byrne, J. and J. Flaherty (2004). Measuring Diversity in Australian Residential Property. Paper presented at the 10th Annual Conference of the Pacific Rim Real Estate Society, January 2004, Bangkok.

27. Caillaud, B. and B. Jullien (2003). Chicken and Egg: Competition among Intermediation Service Providers. *RAND Journal of Economics*, 34(2), 309-328.

28. Carroll, W. (1989). Fixed-percentage commissions and moral hazard in residential real estate brokerage. *The Journal of Real Estate Finance and Economics*, 2(4), 349-365.

29. Case, K.E., Quigley, J.M and R.J. Shiller (2005). Comparing wealth effects: the stock market versus the housing market. *Advances in Macroeconomics*, 5(1).

30. Cervero, R. and M. Duncan (2003). Neighborhood composition and residential land prices: Does exclusion raise or lower values? *Urban Studies*, 41(2), 299-315.

31. Chen, N. (2004). Intra-national versus international trade in the European Union: why do national borders matter?. *Journal of International Economics*, 63(1), 93-118.

32. Cheshire, P. and S. Magrini (2009). Urban growth drivers in a Europe of sticky people and implicit boundaries. *Journal of Economic Geography*, 9, 85-115.

33. Chiswick, B.R. (2008). The Economics of Language: An Introduction and Overview. IZA Discussion Paper No. 3568.

34. Chudik, A. and M.H. Pesaran (2011). Infinite-dimensional VARs and factor models. *Journal of Econometrics*, 163(1), 4-22.

35. Crockett, J.H. (1982). Competition and Efficiency in Transacting: The Case of Residential Real Estate Brokerage. *Journal of the American Real Estate and Urban Economics Association*, 10(2), 209-227.

36. Collins, W.J. and R.A. Margo (2003). Race and the value of owner-occupied housing, 1940-1990. *Regional Science and Urban Economics*, 33(3), 255-286.

37. Coulson, N.E. and M.L. Lahr (2005). Gracing the Land of Elvis and Beale Street: Historic Designation and Property Values in Memphis. *Real Estate Economics*, 33(3), 487-507.

38. Damen, S., Vastmans, F. and E. Buyst (2016). The effect of mortgage interest deduction and mortgage characteristics on house prices. *Journal of Housing Economics*, 34, 15-29.

39. Damiano, E. and H. Li (2007). Price Discrimination and Efficient Matching. *Economic Theory*, 30, 243-263.

40. Damiano, E. and H. Li (2008). Competing Matchmaking. *Journal of the European Economic Association*, 6, 789-818.

41. De Bruyne, K. and J. Van Hove (2013). Explaining the spatial variation in housing prices: an economic geography approach. *Applied Economics*, 45(13), 1673-1689.

42. De Decker, P. (2011). Freiburg of L.A.? *Ruimte & Maatschappij*, 2(4), 1-6.

43. Drukker, D.M., Egger, P. and I.R. Prucha (2013). On two-step estimation of a spatial autoregressive model with autoregressive disturbances and endogenous regressors. *Econometric Reviews*, 32(5-6), 686-733.

44. van Dijk, B., Franses, P. H., Paap, R. and D. van Dijk (2011). Modelling regional house prices. *Applied Economics*, 43(17), 2097-2110.

45. Einav, L. and A. Finkelstein (2011). Selection in Insurance Markets: Theory and Empirics in Pictures. *Journal of Economic Perspectives*, 25(1), 115-138.

46. Einav, L. and J. Levin (2010). Empirical Industrial Organization: A Progress Report. *Journal of Economics Perspectives*, 2010, 24(2), 145-162.

47. Ferreira-Lopes, A. and T.N. Sequeira (2012). Business Cycles Association in a Small Monetary Union: The Case of Switzerland. *Spatial Economic Analysis*, 7(1), 9-30.

48. Fisher, L.M. and A. Yavas (2010). A case for percentage commission contracts: The impact of a "Race" among agents. *The Journal of Real Estate Finance and Economics*, 40(1), 1-13.

49. Florida, R., Mellander, C. and K. Stolarick (2011). Beautiful Places: The Role of Perceived Aesthetic Beauty in Community Satisfaction. *Regional Studies*, 45(1), 33-48.

50. Fotheringham, A.S., Brunsdon, C. and M.E. Charlton (2003). *Geographically weighted regression: the analysis of spatially varying relationships.* John Wiley and Sons, pp. 268.

51. Fuerst, F., McAllister, P. and C.B. Murray (2011). Designer buildings: estimating the economic value of 'signature' architecture. *Environment and Planning A*, 43(1), 166-184.

52. Genesove, D. and C. Mayer (2001). Loss aversion and seller behavior: evidence from housing markets. *Quarterly Journal of Economics*, 2001, 116(4), 1233-1260.

53. Giussani, B. and G. Hadjimatheou (1991). Modeling regional house prices in the United Kingdom. *Papers in Regional Science*, 70(2), 201-219.

54. Glaeser, E.L. and J. Gyourko (2007). Arbitrage in Housing Markets. NBER Working Paper No. 13704.

55. Goffette-Nagot, F., Reginster, I. and I. Thomas (2011). Spatial Analysis of Residential Land Prices in Belgium: Accessibility, Linguistic Border, and Environmental Amenities. *Regional Studies*, 45(9), 1253-1268.

56. Gomes, R. and A. Pavan (2015). Many-to-Many Matching and Price Discrimination. Forthcoming in *Theoretical Economics.*

57. Goodman, A.C. (1978). Hedonic Prices Price Indices and Housing Markets. *Journal of Urban Economics*, 5(4), 471-484.

58. Goos, M., Van Cayseele, P.J.G. and B. Willekens (2011). Platform Pricing in Matching Markets. *Review of Network Economics*, 12(4), 437-457.

59. Gorodnichenko, Y. and L.L. Tesar (2009). Border Effect or Country Effect? Seattle May Not Be so Far from Vancouver After All. *American Economic Journal: Macroeconomics*, 1(1), 219-241.

60. Granger, C.W.J. and J.-L. Lin (1995). Causality in the long run. *Econometric Theory*, 11(3), 530-536.

61. Grinstead, C.M. and J.L. Snell (2012). Introduction to probability. American Mathematical Society.

62. Halleck Vega, S. and P.J. Elhorst (2015). The SLX model. *Journal of Regional Science*, 55(3), 339-363.

63. Hamilton, B.W. (1976). Capitalization of Intrajurisdictional Differences in Local Tax Prices. *American Economic Review*, 66(5), 743-753.

64. Han, L. and S.-H. Hong (2011). Testing cost inefficiency under free entry in the real estate brokerage industry. *Journal of Business and Economic Statistics*, 29(4), 564-578.

65. Han, L. and W.C. Strange (2014). The Microstructure of Housing Markets: Search Bargaining, and Brokerage. In: *Handbook of Regional and Urban Economics*, Vol. 5, Gilles Duranton, Vernon Henderson and William C. Strange (Eds.), Elsevier.

66. Harding, J.P., Rosenthal, S.S. and C.F. Sirmans (2003). Estimating Bargaining Power in the Market for Existing Homes. *Review of Economics and Statistics*, 85(1), 178-188.

67. Haurin, D. (1988). The duration of marketing time of residential housing. *Journal of the American Real Estate and Urban Economics Association*, 16(4), 396-410.

68. Helgers, R. and E. Buyst (2014). Woningprijzen: een regionaal woningprijsmodel. Steunpunt Wonen, Leuven, pp. 51.

69. Helgers, R. and E. Buyst (2016). Spatial and temporal diffusion of housing prices in the presence of a linguistic border: Evidence from Belgium. *Spatial Economic Analysis*, 11(1), 92-122.

70. Hendel, I., Nevo, A. and F. Ortalo-Magné (2009). The Relative Performance of Real Estate Marketing Platforms: MLS versus FSBOMadison.com. *American Economic Review*, 99(5), 1878-1898.

71. Holly, S., Pesaran, M.H. and T. Yamagata (2011). The spatial and temporal diffusion of house prices in the UK. *Journal of Urban Economics*, 69(1), 2-23.

72. Holmes, M.J. and A. Grimes (2008). Is There Long-Run Convergence among Regional House Prices in the UK?. *Urban Studies*, 45(8), 1531-1544.

73. Hsieh, C.-T. and E. Moretti (2003). Can Free Entry Be Inefficient? Fixed Commissions and Social Waste in the Real Estate Industry. *Journal of Political Economy*, 111(5), 1076-1122.

74. Huber, S. and C. Rust (2016). osrmtime: Calculate Travel Time and Distance with OpenStreetMap Data using the Open Source Routing Machine (OSRM). *Stata Journal*, 16(2), 416-423.

75. Iacoviello, M. and S. Neri (2010). Housing Market Spillovers: Evidence from an Estimated DSGE Model. *American Economic Journal: Macroeconomics*, 2(2), 125-160.

76. Ihlanfeldt, K. and T. Mayock (2009). Price discrimination in the housing market. *Journal of Urban Economics*, 66(2), 125-140.

77. Ihlanfeldt, K. and T. Mayock (2012). Information, Search, and House Prices: Revisited. *Journal of Real Estate Finance and Economics*, 44, 90-115.

78. Jacobs-Crisioni, C. and E. Koomen (2014). The influence of national borders on urban development in border regions: An accessibility approach. Unpublished manuscript, VU University Amsterdam.

79. Janssens, R. (2001). Taalgebruik in Brussel. Taalverhoudingen, taalverschuivingen en taalidentiteit in een meertalige stad. Brusselse Thema's 8, VUBPRESS, Brussels, Belgium.

80. Jeziorski, P. (2014). Effects of Mergers in Two-Sided Markets: The U.S. Radio Industry. *American Economic Journal: Microeconomics*, 6(4), 35-73.

81. Kelejian, H.H. and I.R. Prucha (1998). A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances. *Journal of Real Estate Finance and Economics*, 17(1), 99-121.

82. Kim, C.W., Phipps, T.T. and L. Anselin (2003). Measures the benefits of air quality improvement: a spatial hedonic approach. *Journal of Environmental Economics and Management*, 45(1), 24-39.

83. Kuethe, T.H. and V.O. Pede (2011). Regional housing price cycles: a spatio-temporal analysis using US state-level data. *Regional Studies*, 45(5), 563-574.

84. Lambson, V.E., McQueen, G.R. and B.A. Slade (2004). Do Out-of-State Buyers Pay More for Real Estate? An Examination of Anchoring-Induced Bias and Search Costs. *Real Estate Economics*, 32(1), 85-126.

85. Lancaster, K.J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, 74(2), 132-156.

86. Lazrak, F., Nijkamp, P., Rietveld, P. and J. Rouwendal (2014). The market value of cultural heritage in urban areas: an application of spatial hedonic pricing. *Journal of Geographical Systems*, 16(1), 89-114.

87. Lee, R.S. (2013). Vertical Integration and Exclusivity in Platform and Two-Sided Markets. *American Economic Review*, 103(7), 2960 -3000.

88. Leichenko, R., Coulson, E., and D. Listokin (2001). Historic Preservation and Residential Property Values: An Analysis of Texas Cities. *Urban Studies*, 38(11), 1973-1987.

89. Levitt, S. and C. Syverson (2008a). Antitrust implications of outcomes when home sellers use flat-fee real estate agents. *Brookings-Wharton Papers on Urban Affairs*.

90. Levitt, S. and C. Syverson (2008b). Market Distortions When Agents Are Better Informed: The Value of Information in Real Estate Transactions. *Review of Economics and Statistics*, 90(4), 599-611.

91. Lewis, M. (2014). *Flash Boys: a Wall Street Revolt*. New York: W. W. Norton & Company, pp. 288.

92. Listokin, D., Listokin, B., and M. Lahr (1998). The contributions of historic preservation to housing and economic development. *Housing Policy Debate*, 9(3), 431-478.

93. Lindentahl, T. (2016). Value Follows Form: The Shape of Buildings as a Predictor of Property Values. Working Paper.

94. Loertscher, S. and A. Niedermayer (2014). Optimal percentage fees. Working Paper.

95. Loris, I. (2009). Particulier bezit meeste bouwgrond. *Ruimte*, 4, 50-53.

96. Loris, I. (2011). Woonaanbod in Vlaanderen: Analyse van het register van onbebouwde percelen en gemeentelijke woningprogrammaties. *Ruimte & Maatschappij*, 2(4), 72-86.

97. Loyen, R., Buyst, E. and G. Devos (2012). *Struggling for Leadership: Antwerp-Rotterdam Port Competition between 1870-2000*. Heidelberg/New York: Physica-Verlag, pp. 338.

98. Luttmer, E.F.P. (2005). Neighbors as negatives: relative earnings and well-being. *Quarterly Journal of Economics*, 120(3), 963-1002.

99. Luyten, S. and E. Van Hecke (2007). De Belgische stadsgewesten 2001. Statistics Belgium Working Paper.

100. MacDonald, R. and M.P. Taylor (1993). Regional house prices in Britain: long-run relationships and short-run dynamics. *Scottish Journal of Political Economy*, 40(1), 43-55.

101. MacPherson, D.A. and G.S. Sirmans (2001). Neighborhood Diversity and House-Price Appreciation. *Journal of Real Estate Finance and Economics*, 22(1), 81-97.

102. Mahoney, N. and G. Weyl (2014). Imperfect Competition in Selection Markets. NBER Working Paper No. 20411.

103. Malpezzi, S. (2003). Hedonic pricing models: a selective and applied review. In: O'Sullivan, A. and K. Gibb (Eds.), *Housing Economics and Public Policy: Essays in Honor of Duncan Maclennan* (pp. 67-74), Wiley-Blackwell.

104. Mankiw, N.G. and M.D. Whinston (1986). Free Entry and Social Inefficiency. *RAND Journal of Economics*, 17(1), 48-58.

105. Mathys, C. (2012). Economisch belang van de Belgische havens: Vlaamse zeehavens, Luiks havencomplex en haven van Brussel - Verslag 2010. Working Paper 225, Nationale Bank van België (NBB), Brussels, Belgium.

106. McCallum, J. (1995). National Borders Matter: Canada-U.S. Regional Trade Patterns. *American Economic Review*, 85(3), 615-623.

107. Meen, G. (1999). Regional house prices and the ripple effect: a new interpretation. *Housing Studies*, 14(6), 733-753.

108. Meen, G. and M. Andrew (1998). On the aggregate housing market implications of labor market change. *Scottish Journal of Political Economy*, 45(4), 393-419.

109. Melitz, J. (2008). Language and foreign trade. *European Economic Review*, 52(4), 667-699.

110. Merlo, A. and F. Ortalo-Magné (2004). Bargaining over residential real estate: evidence from England" *Journal of Urban Economics*, 56(2),192-216.

111. Micheli, M., Rouwendal, J. and J. Dekkers (2014). Border Effects in House Prices. *Ruhr Economic Papers*, No. 511.

112. Miceli, T.J. (1992). The Welfare Effects of Non-Price Competition among Real Estate Brokers. *Journal of the American Real Estate and Urban Economics Association*, 20(4), 519-532.

113. Miceli, T.J., Pancak, K.A. and C.F. Sirmans (2007). Is the Compensation Model for Real Estate Brokers Obsolete? *Journal of Real Estate Finance and Economics*, 35(1), 7-22.

114. Miller, N.G. (1978). Time on the market and selling price. *American Real Estate and Urban Economics Association Journal*, 6(2), 164-174.

115. Moorhouse, J.C. and M.S. Smith (1994). The Market for Residential Architecture: 19th Century Row Houses in Boston's South End. *Journal of Urban Economics*, 35(3), 267-277.

116. Narwold, A. and J. Sandy (2010). Valuing housing stock diversity. *International Journal of Housing Markets and Analysis*, *3(1), 53-59.*

117. Nationaal Instituut voor de Statistiek (Statistics Belgium) (1954). Algemene volkstelling op 31 december 1947: indeling naar de gesproken landstalen. Ministry of Economic Affairs, Brussels, Belgium.

118. Niedermayer, A. and A. Shneyerov (2014). For-Profit Search Platforms. *International Economic Review*, 55(3), 765-789.

119. Paelinck, J. and L. Klaassen (1979). *Spatial Econometrics*. Farnborough: Saxon House.

120. Palmquist, R.B. (1984). Estimating the Demand for the Characteristics of Housing. *Review of Economics and Statistics*, 66(3), 394-404.

121. Parsley, D.C. and S.-J. Wei (2001). Explaining the Border Effect: The Role of Exchange Rate Variability, Shipping Costs, and Geography. *Journal of International Economics*, 55(1), 87-105.

122. Perloff, J.M., Karp, L.S. and A. Golan (2007). *Estimating Market Power and Strategies*, Cambridge University Press.

123. Pesaran, M.H. (2004). General Diagnostic Tests for Cross Section Dependence in Panels. CESifo Working Paper 1229; IZA Discussion Paper 1240.

124. Pesaran, M.H. (2011). Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure. *Econometrica*, 74(4), 967-1012.

125. Pesaran, M.H. (2007a). A pair-wise approach to testing for output and growth convergence. *Journal of Econometrics*, 138(1), 312-355.

126. Pesaran, M.H. (2007b). A simple panel unit root test in the presence of cross-section dependence. *Journal of Applied Econometrics*, 22(2), 265-312.

127. Pesaran, M.H. and E. Tosetti (2011). Large panels with common factors and spatial autocorrelation. *Journal of Econometrics*, 161(2), 182-202.

128. Quigley, J.M. (1979). What have we learned about urban housing markets?" In: Mieszkowski, P. and M. Straszheim (eds.), Current Issues in Urban Economics, Baltimore: John Hopkins.

129. Quigley, J.M. (2002). Transaction costs and housing markets. Berkeley Program on Housing and Urban Policy, 2002.

130. Rochet, J.-C. and J. Tirole (2003). Platform Competition in Two-Sided Markets. *Journal of the European Economic Association*, 2003, 1, 990-1029.

131. Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1), 34-55.

132. Rouwendal, J. (2007). Mortgage interest deductibility and homeownership in the Netherlands. *Journal of Housing and the Built Environment*, 22(4), 369-382.

133. Rubinstein, A. and A. Wolinsky (1987). Middlemen. *Quarterly Journal of Economics*, 102, 581-594.

134. Rutherford, R.C., Springer, T.M. and A. Yavas (2005). Conflicts between Principals and Agents: Evidence from Residential Brokerage. *Journal of Financial Economics*, 76(3), 627-665.

135. Ryckewaert, M. (2011). *Building the Economic Backbone of the Belgian Welfare State: Infrastructure, planning and architecture 1945-1973*, Rotterdam: 010 Publishers.

136. Rysman, M. (2004). Competition between Networks: A Study of the Market for Yellow Pages. *Review of Economic Studies*, 71(2), 483-512.

137. Schaumans, C. and F. Verboven (2015). Entry and competition in differentiated products markets. *Review of Economics and Statistics*, 97(1), 195-209.

138. Schnare, A. and R.B. Kulick (2009). Do Real Estate Agents Compete on Price? Evidence from Seven Metropolitan Areas. In: *Housing markets and the economy: risk, regulation, and policy*, Edward L. Glaeser and John M. Quigley (Eds.), Cambridge, MA: Lincoln Institute of Land Policy.

139. Schulze, M.-S. and N. Wolf (2009). On the origins of border effects: insights from the Habsburg Empire. *Journal of Economic Geography*, 9(1), 117-136.

140. Simpson, E.H. (1949). Measurement of diversity. *Nature*, 163, 688.

141. Sirmans, G.S., Macpherson, D.A. and E.N. Zietz (2005). The Composition of Hedonic Pricing Models. *Journal of Real Estate Literature*, 13(1), 3-43.

142. Sirmans, C.F. and G.K. Turnbull (1997). Brokerage Pricing Under Competition. *Journal of Urban Economics*, 41(1), 102-117.

143. Spulber, D.F. (1999). Market Microstructure: Intermediaries and the Theory of the Firm. Cambridge: Cambridge University Press.

144. Stein, J.C. (1995). Prices and trading volume in the housing market: a model with down-payment effects. *Quarterly Journal of Economics*, 110(2), 379-406.

145. Stevenson, S. (2004). House price diffusion and inter-regional and cross-border house price dynamics. *Journal of Property Research*, 21(4), 301-320.

146. Stigler, G.J. (1961). The economics of information. *Journal of Political Economy*, 69(3), 213-225.

147. Struyven, D. (2015). Housing Lock: Dutch Evidence on the Impact of Negative Home Equity on Household Mobility. Job Market Paper.

148. Tennekes, J., Harbers, A. and E. Buitelaar (2015). Institutional Arrangements and the Morphology of Residential Development in the Netherlands, Flanders and North Rhine-Westphalia. *European Planning Studies*.

149. Tobler, W.R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46, 234-240.

150. Turnbull, G.K. and C.F. Sirmans (1993). Information, search and house prices. *Regional Science and Urban Economics*, 23(4), 545-557.

151. Turnbull, G.K., Dombrow, J. and C.F. Sirmans (2006). Big House, Little House: Relative Size and Value. *Real Estate Economics*, 34(3), 439-456.

152. US Department of Justice and Federal Trade Commission (2007). Competition in the Real Estate Brokerage Industry.

153. Veblen, T. (1899). *The Theory of the Leisure Class: An Economic Study of Institutions*, New York: Macmillan.

154. Veiga, A. and G. Weyl (2016). Product Design in Selection Markets. Working Paper.

155. Vansteenkiste, I. and P. Hiebert (2011). Do house price developments spillover across euro area countries? Evidence from a global VAR. *Journal of Housing Economics*, 20, 299-314.

156. Verhage, R. (2002). *Local Policy for Housing Development, European Experiences*. Aldershot: Ashgate.

157. Vermeulen, W. and J. Rouwendal (2007). Housing supply and land use regulation in the Netherlands. Tinbergen Institute Discussion Paper 07-058/3, Tinbergen Instititute.

158. Watkins, C. (1998). Are new entrants to the residential property market informationally disadvantaged? *Journal of Property Research*, 15(1), 57-70.

159. Weyl, G. (2010). A Price Theory of Multi-Sided Platforms. *American Economic Review*, 1642-1672.

160. Weyl, G. and M. Fabinger (2013). Pass-Through as an Economic Tool: Principle of Incidence under Imperfect Competition. *Journal of Political Economy*, 121(3), 528-583.

161. White, A. and G. Weyl (2015). Insulated Platform Competition. Working Paper.

162. Wiley, J.A., Benefield, J.D., and M.T. Allen (2012). Cyclical Determinants of Brokerage Commission Rates. *Journal of Real Estate Finance and Economics*, 48(1), 1-24.

163. Willekens, B., Helgers, R., Goos, M. and E. Buyst (2015). Intermediation in Markets with Buyer and Seller Selection: Theory and an Application to Real Estate Brokerage. CES Discussion Paper 15.32.

164. Wu, D.-M. (1973). Alternative tests of independence between stochastic regressions and disturbances. *Econometrica*, 41(4), 733-750.

165. Yavas, A. (1994). Middlemen in Bilateral Search Markets. *Journal of Labor Economics*, 12, 406-429.

166. Yavas, A. (2001). Impossibility of a Competitive Equilibrium in the Real Estate Brokerage Industry. *Journal of Real Estate Research*, 21(3), 187-200.

167. Yinger, J. (1981). A Search Model of Real Estate Broker Behavior. *American Economic Review*, 71(4), 591-604.

168. Zahirovic-Herbert, V. and S. Chatterjee (2012). Historic Preservation and Residential Property Values: Evidence from Quantile Regression. *Urban Studies*, 49(2), 369-382.

169. Zietz, E.N. and G.S. Sirmans (2011). Real Estate Brokerage Research in the New Millennium. *Journal of Real Estate Literature*, 19(1), 5-40.

Doctoral dissertations from the Faculty of Economics and Business, see:
http://www.kuleuven.be/doctoraatsverdediging/archief.htm